

Best Available Copy

# SEARCH REQUEST FORM

11-38

Requestor's  
Name: \_\_\_\_\_

Serial  
Number: \_\_\_\_\_

Date: \_\_\_\_\_

Phone: \_\_\_\_\_

Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

FOR OFFICIAL USE ONLY

## STAFF USE ONLY

Date completed: 11-10-97

Searcher: MAR 16

Terminal time: 6

Elapsed time: 10

CPU time: \_\_\_\_\_

Total time: 16

Number of Searches: 1

Number of Databases: 11

### Search Site

☐ STIC

☒ CM-1

☐ Pre-S

### Type of Search

☒ N.A. Sequence

☒ A.A. Sequence

☐ Structure

☐ Bibliographic

### Vendors

☒ IG-MPRCH

☐ STN

☐ Dialog

☐ APS

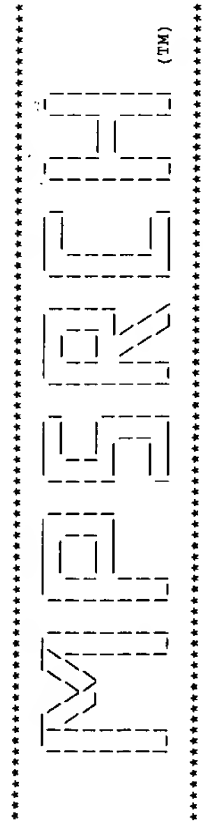
☐ Geninfo

☐ SDC

☐ DARC/Questel

☐ Other

**This Page Blank (uspto)**



\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
MPsrch\_n n.a. n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 10:15:18 1997; Maspar time 820.34 Seconds  
Tabular output not generated.  
798.029 Million cell updates/sec  
Title: >US-08-842-827-1  
Description: (1-1563) from US08842827.seq  
Perfect Score: 1563  
N.A. Sequence: 1 CCTGTGCGAGACGCCGCGG.....CCAAAAA.....AAAAA 1563  
Comp: GGACACCTCTCTCCGCCGCC.....GGTTTTTTTTTTTTTTTT

Scoring table:  
Gap 6  
Nmatch STD : Dbase 0: Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database:  
EST-STS  
1: EST2: EST3: EST4: EST5: EST6: EST7: EST8: EST9: EST10: EST11: EST12: EST13: EST14: EST15: EST16: EST17: EST18: EST19: EST20: EST21: EST22: EST23: EST24: EST25: EST26: EST27: EST28: EST29: EST30: EST31: EST32: EST33: EST34: EST35: EST36: EST37: EST38: EST39: EST40: EST41: EST42: EST43: EST44: EST45: EST46: EST47: EST48: EST49: EST50: EST51: EST52: EST53: EST54: EST55: EST56: EST57: EST58: EST59: EST60: EST61: EST62: EST63: EST64: EST65: EST66: EST67: EST68: EST69: EST70: EST71: EST72: EST73: EST74: EST75: EST76: EST77: EST78: EST79: EST80: EST81: EST82: EST83: EST84: EST85: EST86: EST87: EST88: EST89: EST90: EST91: EST92: EST93: EST94: EST95: EST96: EST97: EST98: EST99  
EST-STS-TWO  
100: EST100: 101: EST101: 102: EST102: 103: EST103: 104: EST104: 105: EST105: 106: EST106: 107: EST107: 108: EST108: 109: EST109: 110: EST110: 111: EST111: 112: EST112: 113: EST113: 114: EST114: 115: EST115: 116: EST116: 117: EST117: 118: EST118: 119: EST119: 120: EST120: 121: EST121: 122: EST122: 123: EST123: 124: EST124: 125: EST125: 126: EST126: 127: EST127: 128: EST128: 129: EST129: 130: EST130: 131: EST131: 132: EST132: 133: EST133: 134: EST134: 135: EST135: 136: EST136: 137: EST137: 138: EST138: 139: EST139: 140: EST140: 141: EST141: 142: EST142: 143: EST143: 144: EST144: 145: EST145: 146: EST146: 147: EST147: 148: EST148: 149: EST149: 150: EST150: 151: EST151: 152: EST152: 153: EST153: 154: EST154: 155: EST155: 156: EST156: 157: EST157: 158: EST158: 159: EST159: 160: EST160: 161: EST161: 162: EST162: 163: EST163: 164: EST164: 165: EST165: 166: EST166: 167: EST167: 168: EST168: 169: EST169: 170: EST170: 171: EST171: 172: EST172: 173: EST173: 174: EST174

Statistics: Mean 11.804; Variance 2.893; scale 4.080

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
C 1	528	33.8	587 75	H97570	yx03b06.s1 Homo sapie	0.00e+00
C 2	515	32.9	545 137	AA043085	zk48b02.r1 Soares pre	0.00e+00
C 3	507	32.4	552 159	AA036943	zk30f04.s1 Soares pre	0.00e+00
C 4	501	32.1	546 171	W67667	zk38a11.s1 Soares fet	0.00e+00
C 5	461	29.5	465 160	AA040858	zk48b02.s1 Soares pre	0.00e+00
C 6	448	28.7	472 114	W04968	za43g09.r1 Soares fet	0.00e+00
C 7	414	26.5	490 92	N31047	yx51b07.r1 Homo sapie	0.00e+00
C 8	406	26.0	446 171	W67666	zk38a11.r1 Soares fet	0.00e+00
C 9	406	26.0	474 113	W01275	za40f11.r1 Soares fet	0.00e+00
C 10	405	25.9	456 169	W45106	zc21a10.r1 Soares sen	0.00e+00
C 11	395	25.3	422 169	W45155	zc21a10.s1 Soares sen	0.00e+00
C 12	381	24.4	456 67	H68363	yx82f11.r1 Homo sapie	0.00e+00
C 13	380	24.3	471 189	AA152123	z141g02.r1 Soares pre	0.00e+00
C 14	376	24.1	415 151	W24605	zb63f08.r1 Soares fet	0.00e+00
C 15	377	24.1	479 74	H90961	yu86e01.r1 Homo sapie	0.00e+00
C 16	374	23.9	423 78	R97295	Yq74806.r1 Homo sapie	0.00e+00
C 17	367	22.8	452 85	H54373	Yq33g09.r1 Homo sapie	0.00e+00
C 18	357	22.8	471 85	H57543	Yq05d10.r1 Homo sapie	0.00e+00
C 19	357	22.8	480 169	W45113	zc21c09.r1 Soares sen	0.00e+00
C 20	354	22.6	404 62	H17855	ym36b12.s1 Homo sapie	0.00e+00
C 21	347	22.2	438 4	T70188	Yc18c09.s1 Homo sapie	0.00e+00
C 22	347	22.2	473 85	H54296	Yq33g09.s1 Homo sapie	0.00e+00
C 23	342	21.9	356 165	C17623	Human placenta cDNA 5	0.00e+00
C 24	343	21.9	445 85	H57544	Yr05d10.s1 Homo sapie	0.00e+00
C 25	334	21.4	386 5	T70311	Yd26h07.s1 Homo sapie	0.00e+00
C 26	322	20.6	331 158	AA033777	zk19g11.r1 Soares pre	0.00e+00
C 27	321	20.5	351 85	H57213	Yr08d07.r1 Homo sapie	0.00e+00
C 28	319	20.4	332 158	AA033778	zk19g11.s1 Soares pre	0.00e+00
C 29	315	20.2	408 7	T80888	Yd26h07.r1 Homo sapie	0.00e+00
C 30	314	20.1	357 159	AA037575	zk34f08.r1 Soares pre	0.00e+00
C 31	310	19.8	357 64	H26307	Y151f02.s1 Homo sapie	0.00e+00
C 32	308	19.7	386 159	AA036941	zk30f04.r1 Soares pre	0.00e+00
C 33	304	19.4	413 67	H68384	Yr82f11.s1 Homo sapie	0.00e+00
C 34	298	19.1	422 4	T69888	Yc18c09.r1 Homo sapie	0.00e+00
C 35	287	18.4	304 198	H96367	Yw61f11.s1 Soares pre	0.00e+00
C 36	270	17.3	279 108	HSC37E112	H. sapiens partial cd	0.00e+00
C 37	269	17.2	315 9	T90107	Yd33g02.s1 Homo sapie	0.00e+00
C 38	267	17.1	287 142	N86994	L1773f fetal heart, L	0.00e+00
C 39	257	16.4	351 36	R07363	Yq96f08.s1 Homo sapie	0.00e+00
C 40	249	15.9	352 189	AA150023	z141g01.s1 Soares pre	0.00e+00
C 41	248	15.9	331 37	R09946	Yf30f03.r1 Homo sapie	0.00e+00
C 42	247	15.8	325 128	W92816	zh89g08.s1 Soares fet	0.00e+00
C 43	237	15.2	247 108	HSC36C022	H. sapiens partial cd	0.00e+00
C 44	234	15.0	511 154	AA008988	mg99e04.r1 Soares mou	0.00e+00
C 45	226	14.5	354 73	H90307	yu86e01.s1 Homo sapie	0.00e+00

ALIGNMENTS

1 H97570 587 bp mRNA EST 12-DEC-1995  
LOCUS Yx03b06.s1 Homo sapiens cDNA clone 260627 3'  
DEFINITION H97570  
ACCESSION g1118455  
NID EST  
KEYWORDS human clone-260627 primer-m13 -40 forward library=Soares melanocyte  
SOURCE

2NHM vector-pT7T3D (Pharmacia) with a modified polylinker host-strIOB (ampicillin resistant) SalI-Not I Rsite2-Eco RI Male, 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCGAATGGAGCGCGGAGTGTGTGTGTGTGTGT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fátima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM	REFERENCE
Homo sapiens	1 (cases 1 to 587)
Euarchontae	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Hollman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
Metazoa	The WashU-Merck EST Project
Unpublished (1995)	

**CONTACT:** Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 351  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Location/Qualifiers

Source	1..587	/organism="Homo sapiens"			
		/clone="260627"			
mRNA	<1..>587				
BASE COUNT	173 a	117 c	120 g	174 t	3 others
ORIGIN					
Query Match	33.8%	Score 528;	DB 75;	Length 587;	
Best Local Similarity	96.6%	Pred. No. 0.00e+00;			
Matches	565;	Conservative	0;	Mismatches 16;	Indels 4;
					Gaps 4;

Db	1	ggtctcccatcacattttaatgtatgtattattttaataatacaacatcatttcacagaagca	50
Cp	1545	ggtctcccatcacattttaatgtatgtattattttaataatacaacatcatttcacagaagca	1486
Db	61	tattacatacattgtttatatacaataagcattcacatttttttataaaaaatgtatacaggtgg	120
Cp	1485	tattacatacattgtttatatacaataagcattcacatttttttataaaaaatgtatatacaggtgg	1426
Db	121	ggcactgttttggtagaagccttgaggtttttttaaagatttagcgtattagataaac	180
Cp	1425	ggcactgttttggtagaagccttgaggtttttttaaagatttagcgtattagataaac	1366
Db	181	actgagtttaaaagtaaacatgtatcacacaaagtg-gcatccaagagggcatacagcagca	239
Cp	1365	actgagtttaaaagtaaacatgtatcacacaaagtg-gcatccaagagggcatacagcagca	1306
Db	240	gaagctctttaagccttgatcacaccaggaagaagatcatcctcttgccttgggaaatc	299
Cp	1305	gaagctctttaaaagccttgatcacaccaggaagaagatcatcctcttgccttggcgaatc	1246
Db	300	attttccctttagaanaacaggccagcttcaacctgggaccacctgctgccttcaaggctggg	359
Cp	1245	attttcccttttgaanaacaggccagcttcaacctgggaccacctgctgccttcaaggctggg	1187
Db	360	tgattgctcggaatagtgattcccaagttgttggtgtttcatgcagagatttgtatgagatcc	419

Accession	Sequence	Position
Cp	1186 TGATTGCTGGATAGTGATTCCAGTGTGTGGTGTTCATGTCAGAGATTGTATGAGAGTCC	1127
Db	420 tctctcttctctttttaaagaagttcttttccctgaagaaatccgatacatatacagca	479
Cp	1126 TCGCTCTTTTCTTTTAAAGAGGTTCTTCTTTTGAGGAANTCCGATACATATACACA	1067
Db	480 actaatattgcaacagagctcctcgaatgagtcagtcacacacatcgctccagtgagg	539
Cp	1066 ACTAATATGCAACAGAGTCCCTGATGAGTCCAGTCACACATCGCTCC - AGTGGTG	1008
Db	540 gtttaataatccgmactccgngaaggncccatataatagggaagcg	584
Cp	1007 -TTTATATACAGAACTTCAGAAAGGCCACATAAATGGATAGCG	964

RESULT	AA043085	545 bp	mRNA	EST	04-SEP-1996
LOCUS	2X84B02.1	Soares pregnant uterus	NbHPU	Homo sapiens	CDNA clone
DEFINITION	485027.5.				
ACCESSION	AA043085				
NID	91520979				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;				
	Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
REFERENCE	1 (bases 1 to 545)				
AUTHORS	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.				
TITLE	The WashU-Merck EST Project				
JOURNAL	Unpublished (1995)				

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 485.

FEATURES

1...545

Location/Qualifiers

/organism="Homo sapiens"

/notes="Organ:uterus; Vector: p7T73-Pac; Site\_1: Not I  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I  
oligo(dT) primer [5',  
AAGTGAACAATTCGCGCGCCCTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the N  
and Eco RI sites of the modified p7T73 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."

	159 a	116 c	105 g	153 t	2 others
mRNA					
BASE COUNT					
ORIGIN					
	/clone_48027/				
	/clone_lib=Soares pregnant uterus NBHPU"				
	/sex="female"				
	/dev_stage="adult"				
	/lab_host="DH10B"				
	<1..>545				

```

Query Match      32.9%; Score 515; DB 137; Length 545;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;
Matches 540; Conservative 0; Mismatches 3; Indels 4; Gaps
Db 1 cccacactgcattgtgttgcgtatccattatgtgggccttctcgnattct 60
      |||||

```



```
QY 939 CCACACTGCAATTTGGTCTTGTCGCCGTATCCATTATGTGGCCCTTTCTCGAGTTTCT 998
Db 61 gattataaacacactggagagatgttgactgactgaactattcaggagctctgttgca 120
|||||
QY 999 GATTATAAACACACTGGAGGATGTGTGACTCGGATCTCAATCAGGAGCTCGGTGCA 1058
Db 121 atattagttcgtgtatgtatcgatctcttctcaaaagaagaactcttttaagaaga 180
|||||
QY 1059 ATATTAGTTGCTGTATGTATCGGATTTCTTCAAGAAAGAACTCTTTTAAAGAA 1118
Db 181 aaagaggagactctcatcaactctgcatgaacacacacacactgggaatcactatcg 240
|||||
QY 1119 AAGAGAGGACTCTATACAACTCTGATGAACACACCACTGGGAATCACTATCGG 1178
Db 241 agcaatcacacagccttgaagcagcagggcgccaggtgaaagctgacctgttttctaa 300
|||||
QY 1179 AGCAATCACAGCCTTGAAGGACAGAGGCGCCAGGTGAAGCTGSCCTGTTTCTAAA 1238
Db 301 gaaaaatgattcccaagp-aagagatgcatcttctctctgggtgataagcccttaa 359
|||||
QY 1239 GGAATGATGTCACAGGCAAGGAGGATGATCTTTCTCTGGTGTACAGCCCTTAA 1298
Db 360 agactctgctgctatgctctgctgctgacacatttgggtacatagttacattta 419
|||||
QY 1299 AGACTTCTGCTGATATGCTCTTGGATGCACACATTTGTGTACATAGTTACCTT 1358
Db 420 actcaatggttatctaagctctaaactcatttaaaaaactcccaagcctccacaaa 479
|||||
QY 1359 ACTCAGTGGTATCTAATAA-GCTCTAAACTCATTAATAAAACCTCCAGCCCTCCACAA 1417
Db 480 acsagtccccacgtgtatcatctttttataaaaaa-tgtaatgcttatgtataa-catgt 537
|||||
QY 1418 ACAGTCCCCACCTGATACATTTTATTAAAAAATGTAATGCTTATGTATAACATGT 1477
Db 538 atgtaat 544
|||||
QY 1478 ATGTAAT 1484

RESULT 3
LOCUS AA036943 552 bp mRNA EST 26-AUG-1996
DEFINITION 2k30f04.s1 Soares pregnant uterus NbHPU Homo sapiens CDNA clone
ACCSSION AA036943
NID 91510000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Hillier,L., Clark,M., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 452.
LOCATION/Qualifiers
FEATURES
source 1..552
/organism="Homo sapiens"
```

```
/note="Organ: uterus; Vector: pT7t3-Pac; Site:1; Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(5') primer [5'
AACGGAAGAATTCGGCGCCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7t3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/clone="47203"
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(<1..>552)
BASE COUNT 158 a 105 c 110 g 177 t 2 others
ORIGIN

Query Match 32.4%; Score 507; DB 159; Length 552;
Best Local Similarity 98.5%; Pred. No. 0.00e+00;
Matches 533; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Db 8 ttgggtctcccatcacatttaataatgtattatattttaaatacaacatcatcattacagaa 67
|||||
Cp 1548 TTGGTCTCCCATACATTTTAATATGTATATATTTAAATCAAAATCATTCATAGAA 1489
Db 68 gatattacatacatgtttatatacagaatcatttttttaataaaaaatgtatacagg 127
|||||
Cp 1488 GCATATTACATACATGTTTATACATAAGCATTTACATTTTAAATAAAATGTATACAG 1429
Db 128 tggggcacctgttttggtagagggcttgagggttttttaagtgagtttagagcattagata 187
|||||
Cp 1428 TGGGGCACCTGTTTGGTGAAGGCTTGGAGTGTTTTAAATGAGCTTTAGAGCTATTAGATA 1369
Db 188 accactgagttaaaggtaactatgtacacacaaagtg-gcatccaagaggcatagcagca 246
|||||
Cp 1368 ACCACTGAGTTAAAGGTAACATATGTACACACAAAGTGTGCATCCAAAGAGGCATATACGA 1309
Db 247 gcagaagctcttaaaaggctgtacaccaggagaagaatgcattcctctgtcctgttgn-a 305
|||||
Cp 1308 GCAGAGTCTTTAAAGGCTTGTACACCAGGAAGAAAGATGCATCTCTTGGCTTGGGCA 1249
Db 306 atcatttctctttagaaaaacaggcagcttcaacctgggaccctgtgctttccaagct 365
|||||
Cp 1248 ATCATTTTCTTTTAGAAAACAGGCGAGCTTCACTGGGACCCCTGCTGCTTCAAGGCT 1189
Db 366 ggtgattgctggatagtgattcccgatgctgttgggtgttttcacgcagagtgatagag 425
|||||
Cp 1188 GGTGATTGCTCGGATAGTATCCAGTT-GTTGGTGTTCATGCAGAGTGTATGAGAG 1130
Db 426 tctctcttcttcttctttaaagaaggtcttctcttgaagaataccatacata 485
|||||
Cp 1129 TCCTCTCTTTTCTTCTTTTAAAGAAAGTCTTTCTTTTGAAGAAATCCGATACATATACA 1070
Db 486 gccactaatattgcaaccgagctccctgggaatagtcacagtcacacacatcgtccagt 545
|||||
Cp 1069 GCACTTAATATTGCAACACAGAGCTCCCTG-AATGAGTCCAGTCAACACATCGCTCCAGTG 1011
Db 546 g 546
Cp 1010 g 1010

RESULT 4
LOCUS W67667 546 bp mRNA EST 15-OCT-1996
DEFINITION zd38a11.s1 Soares fetal heart NBHH19W Homo sapiens CDNA clone
ACCSSION W67667
NID 91376762
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
```





Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

## TITLE

## JOURNAL

## COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

High quality sequence stops: 385

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

## FEATURES

Location/Qualifiers  
1..490  
/organism="Homo sapiens"  
/clone="265309"

## BASE COUNT

ORIGIN  
131 a 107 c 111 g 138 t 3 others

Query Match 26.5%; Score 414; DB 97; Length 490;

Best Local Similarity 97.1%; Pred. No. 0.00e+00;  
Matches 476; Conservative 0; Mismatches 4; Indels 10; Gaps 10;

Db 1 gngcctcaactcttggatgttgcgtccagattggtcaaaatacaactgcagcgatgg 60  
QY 719 GCGGCGCTCACTTCTGGATGTTGTGATCCAGATTGGTCAAAAATCAACTGCAGCGATGG 778  
Db 61 ttacattgaactacatcatgtcaggggaatgcagaaanagtttaaggaagcagatttc 120  
QY 779 TTACATTGAATACATATGTCGAGGGAATGCAGAAAGATTAAGGAAGCAGGTTGTC 838  
Db 121 cttctattcaggccactcttctgtttccatgtactgcgtctgttggcactttatct 180  
QY 839 CTTCTATTTCAGGCCACTCTTCGTTTTCATGTACTGTCGATGCTGTTTGTGGCACTTTATCT 898  
Db 181 tcaagccaggatgaaggagactggcgaagactcttaccccaactgcactgaatttgctc 240  
QY 899 TCAAGCCAGGATGAGGAGACTGGGCAAGACTCTTACGCCCCACACTGCCAATTGGTCT 958  
Db 241 ttttccgtatccatttatgtgggccttctcagttctcagttcttgattataaacaaccactggag 300  
QY 959 TGTTCGGCTATCCATTATTATGTGGCCCTTCTCGAGTTTCTGATTATATAACACCACTGGAG 1018  
Db 301 cgatgtgttgactgactcattccaggagctctcgtgtgcaatattagttcgtgtatatgt 360  
QY 1019 CGATGCTGTGACTGGACTATTCAGGAGCTCTGTTGCAATATTAGTTCGTATATGT 1078  
Db 361 atcggattcttcaagaagaactctctttaaagaagaagaaggagagactctcca 420  
QY 1079 ATCGGATTCTCTCAAGAAGAAGCTTC-TTTTAAAGAAAGAAAG-AGGAGGACTCTC-A 1135  
Db 421 tacaactctgcctggaacacacacaaactggggaatcactaaatccagagcaatccacc 480  
QY 1136 TACAACCTC-TGC-ATG-AAACCCACAACTGSGAAT-CACTA-TCCG-AGCAATC-ACC 1188  
Db 481 agccttgaag 490  
QY 1189 AGCCTTGAA 1198

RESULT 8 W67666 446 bp mRNA EST 15-OCT-1996  
LOCUS 3438a11.r1 Soares fetal heart NBH19W Homo sapiens cDNA clone  
DEFINITION 342908 5'.  
ACCESSION W67666  
NID 91376761  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens

## REFERENCE

## AUTHORS

1 (bases 1 to 446)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maier, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

## TITLE

## JOURNAL

## COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 634 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 328.

## FEATURES

## source

Location/Qualifiers

1..446

/organism="Homo sapiens"

/note="Organ: heart; Vector: pT73D (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer (5'

TGTTCAATCTGAAGTGGAGCGCGCATCTTTTTTTTTTTT 3'),

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by

M.Fatima Bonaldo. This library was constructed from the

same fetus as the fetal lung library, Soares fetal lung

NbH19W."

/clone="342908"

/clone\_lib="Soares fetal heart NBH19W"

/sex="unknown"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

<1..>446

## BASE COUNT

ORIGIN  
122 a 102 c 100 g 121 t 1 others

## Query Match

Best Local Similarity 26.0%; Score 406; DB 171; Length 446;

Matches 431; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

Db 2 ttatctcgaagcagatgaaggagactgggcaagactcttacgccccacactgcaat 61

QY 892 TTTATCTCAAGCCAGGATGAAGGAGACTGGCAAGACTCTTACGCCCCACACTGCAAT 951

Db 62 ttgctctgttcggtatccattatgtgggcctttctcgagttctgtattataaaccc 121

QY 952 TTGGTCCTGTGTTCGGATATCCATTATGTGGGCTTTCGAGTTTCTGATTATAAGACC 1011

Db 122 actggagcgatgtgtgactggactcattcaggagctctgtgtgcaatattagttgctg 181

QY 1012 ACTGGAGCGATGTGTTGACTGGACTCATTCAGGAGCTCTGTTGCAATATTAGTTGCTG 1071

Db 182 tatatgacgattcttccaaagaagactcttttaagaagaagaagagagagact 241

QY 1072 TATATGATCGGATTTCTTCAAGAAGAAGACTCTCTTTAAAGAAGAAAGAGGAGGACT 1131

Db 242 ctacataactctgcataaaccacacacactgggaatcactatcccgagcaatccaccgc 301

QY 1132 CTCATACAACCTCTGCATGAACACCAACCACTGGGAATCACTATCCGACCAATCACCAGC 1191

Db 302 cttgaaggcagcagggtgcc-agggaagcggcctgtttcttaaggaaaaatgattgc 360











FEATURES IMAGE Consortium (info@image.llnl.gov) for further information.

Source 1..479  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="240696"  
<1..>479

BASE COUNT 131 a 99 c 102 g 142 t 5 others

ORIGIN

Query Match 24.1%; Score 377; DB 74; Length 479;  
Best Local Similarity 97.9%; Pred. No. 0.00e+00;  
Matches 420; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

Db 1 atttggtctnngtgcgctatccattatgtggccttctcgtgagttctgattataaaca 60  
QY 950 ATTGGTCTTGTCGCGATCCATTATGTGGCCTTCTCGAGTTCTGATTATAAACA 1009

Db 61 ccactggagcagatgtgtgactgactcattcaggagctcgtgtgcaattattgttc 120  
QY 1010 CCACTGGAGCGATGTGTGACTGACTCATTCAGGAGCTCTGTTGCAATTATTAGTTGC 1069

Db 121 tgtatatgtatggatttcttcaagaagaactcttttaagaagaagaaagaggagga 180  
QY 1070 TGTATATGTATCGGATTCTTCAAGAAAGAACTCTTTTAAAGAAAGAAAGAGAGGGA 1129

Db 181 ctctcatcaactctgcatgaacacacacactgggaatcactatccgagcaatcacca 240  
QY 1130 CTCTCATCAACTCTGCAATGAACACCAACCAACTGGGAATCACTATCCGAGCAATCACCA 1189

Db 241 gccttgaagcagcaggggtgccaggtgagcgtggcctgtttctaaagggaataatgatt 300  
QY 1190 GCCTTGAAGGCAGCAGGGTGCCCGAGGTGAAGCTGGCCTGTTTCTAAAGGAAATGATT 1249

Db 301 gccacaaggcagagagatgcattcttctcctgggtgtacaagcctttaagaactttcg 360  
QY 1250 GCCACAAGGCAAG-AGGATGCATCTTCTCCTGG-TGTACAAGCCTTTAAAGACTTCTG 1307

Db 361 ctgctgctatgcctcttgggagggcacactttgtgtgtnacatagtttacctttnaactc 420  
QY 1308 CTGCTGATATGCTCTTGG-ATG-CACACTTTGTGTGT-ACATAGTT-ACCTTT-AACTC 1362

Db 421 agtgattat 429  
QY 1363 AGTGGTAT 1371

Search completed: Thu Nov 6 10:31:38 1997  
Job time : 980 secs.

**This Page Blank (uspto)**

\*\*\*\*\*

MP5RELH

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MP5rch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 09:39:11 1997; Maspar time 1301.12 Seconds  
Tabular output not generated. 1319.331 Million cell updates/sec

Title: >US-08-842-827-1  
Description: (1-1563) from US08842827.seq  
Perfect Score: 1563  
N.A. Sequence: 1 CCTGTGGGAGAGCGCGG.....CCAAAAA.....AAAAA 1563  
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-new3

1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database:

genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL  
Database: u-emb150\_99  
122: part1

Statistics: Mean 11.930; Variance 6.210; scale 1.921

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	577	36.9	1212	85	D84376	House mouse; Musculus	0.00e+00
2	413	26.4	1490	88	MUSHP1P	Mus musculus (clone H	8.00e-297
3	130	8.3	3353	82	HUMORFA02	Human mRNA for KIAA00	1.11e-71
4	111	7.1	2206	91	RNDRI42	R.norvegicus mRNA for	2.52e-57
5	102	6.5	1444	78	HSU79294	Human clone 23748 mRN	1.29e-50
6	41	2.6	215	57	128278	Sequence 5 from paten	2.60e-08
7	34	2.2	215	57	128278	Sequence 5 from paten	3.22e-04
8	32	2.0	1727	45	SSM2RR	S.solidissima mRNA fo	4.10e-03
9	31	2.0	1460	40	DMG11584	Drosophila melanogast	1.42e-02
10	31	2.0	14626	35	BNOF1CEG	Bombax mori fibroin 1	1.42e-02
11	31	2.0	19517	40	DMU37541	Drosophila melanogast	1.42e-02
12	30	1.9	1427	39	DDAAC3	Dictyostelium discoide	4.81e-02
13	29	1.9	1512	46	BBCA	B.taurus mRNA for cyc	1.59e-01
14	29	1.9	1837	91	RNP2XMR	R.norvegicus P2X mRNA	1.59e-01
15	29	1.9	1862	64	PSMANC	Plomycus sp. mRNA fo	4.81e-02
16	29	1.9	8567	16	CAGNAT2	C.aureus GNAT2 gene	1.59e-01
17	28	1.8	363	91	RNA7CINT	R.norvegicus mRNA for	5.15e-01
18	28	1.8	417	40	DOATRICH	D.obscura A-T-rich re	5.15e-01
19	28	1.8	793	122	MDA430	Mouse Murri mRNA, exo	5.15e-01
20	28	1.8	793	118	D85430	Mouse Murri mRNA, exo	5.15e-01
21	28	1.8	793	14	MDA430	Mouse Murri mRNA, exo	5.15e-01
22	28	1.8	1524	43	LPACT3	L.polyphemus mRNA for	5.15e-01
23	28	1.8	1737	39	DD19	Dictyostelium discoid	5.15e-01
24	28	1.8	2049	44	S67872	Zm-glucose-6-phosphat	5.15e-01
25	28	1.8	2630	43	PCU53325	Plasmodium chabaudi d	5.15e-01
26	28	1.8	3077	44	PFU07706	Plasmodium falciparum	5.15e-01
27	28	1.8	3774	91	RNIAC	R.norvegicus mRNA for	5.15e-01
28	28	1.8	6372	39	DDICNPA	Dictyoselium discoid	5.15e-01
29	28	1.8	10359	44	PFARPI	Plasmodium falciparum	5.15e-01
30	28	1.8	192274	110	HS435C23	Human DNA sequence **	5.15e-01
31	28	1.8	192274	5	HS435C23	Human DNA sequence **	5.15e-01
32	27	1.7	347	54	A23327	O.sativa mRNA for T23	1.62e+00
33	27	1.7	451	60	BNUS3459	Brassica napus antif	1.62e+00
34	27	1.7	825	65	SCCOX1	Yeast cytochrome c ox	1.62e+00
35	27	1.7	1095	61	DDICSA	Dictyostelium discoid	1.62e+00
36	27	1.7	1128	43	MPUI5677	Myrmecia pilosula H18	1.62e+00
37	27	1.7	1381	39	C7HRP45MR	C.tentans mRNA for hr	1.62e+00
38	27	1.7	1702	43	PFARGPTA	Plasmodium falciparum	1.62e+00
39	27	1.7	1803	73	HSARGBPTA	H.sapiens mRNA for Ar	1.62e+00
40	27	1.7	2035	39	DDU53884	Dictyostelium discoid	1.62e+00
41	27	1.7	2045	44	PNLGRSYN	Panulirus argus gluta	1.62e+00
42	27	1.7	2798	42	HYDHP701B	Hydra magnipapillata	1.62e+00
43	27	1.7	3700	39	DDGP80G	D.discoidium gp80 gen	1.62e+00
44	27	1.7	7989	44	PFU31083	Plasmodium falciparum	1.62e+00
45	27	1.7	9353	44	PFU67959	Plasmodium falciparum	1.62e+00

ALIGNMENTS

RESULT	1	D84376	1212 bp	mRNA	ROD	03-OCT-1996
LOCUS		House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid phosphatase, complete cds.				
DEFINITION		D84376				
ACCESSION		g1487872				
NID		Phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.				
KEYWORDS		Mus musculus kidney cDNA to mRNA.				
SOURCE		Mus musculus				
ORGANISM		Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE		1 (bases 1 to 1212)				
AUTHORS		Kanoh,H.				
TITLE		Direct Submission				
JOURNAL		Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Hideo Kanoh, Sapporo Medical University School of Medicine, Department of Biochemistry, West-17, Sapporo, Hokkaido 060, Japan (E-mail:kanoh@serpent.cc.sapmed.ac.jp, Tel:011-611-2111(ex.2290), Fax:011-612-5861)				



```

/codon_start-1
/db_xref="pid:g473933"
/translation="LLPKMADAFGDELFSVEGDSSTTAAGTKDKKDKGKWKGPSPG
ADKAGKREFGDKQSESTNKNKRDVDFEGTEIFGKPKRIEESITEDLSLADLMP
VKQSVETGCGTHEVALPAEDYIPLKPRVGAKEEYPTLDAFOREAIOCVDNQNS
VLASNPASGKTCVCAEYALALREKQVIFETSPKALSNOKYEMSEEGVQWGLTIL
DVINPSPACLVNTTILSLMLXRGSEVMEVAVIDEIHYMRDSRGVQVMEGLTIL
LPDNVHYVLSAIPNARQFAEWCHLKHQCPCHVYITDYRTPLQHIITFPAGDGLHL
VVDENGDFREDNNTANQVLRDGLAKDQKRGKTKGPSNFKVKIMIMERNQFP
VIIFSPSKRCKEAYALQMTKLDFTNDEKKMVEEVSNAIDCLSDDEKDKLPQVEHVL
LLKRGITIHGGGLLPILKETIILFSEGLIKALFATETFANGINPPTVLTNARFK
DKGDFRISSEGYIQSGAGRRGMDRGIVILAWDEKMSPTIGQLKLGASDPLNSA
FHTYNNLLRVEINEPTMELKSFQFOHYRAIPGVVEKVNSEQNKIVIPNE
ESVVIYKIQOOLAKIGKIEEYIHKPKYCFPLQPGRLVKVKNEGDDFGVGVVNF
KKSNNPNSGELDPLVYVEVILRCSEKSLKNSATEAKPKADEKGEWQVPLVHLI
SAISSVRLYIPKDLRPVDNRQSVLSIQEVQKRPEDGIPLLDPTDDMGIOQGLKVKI
QKVEAFPHRMVSHPLNDNENFVILCEKKNSTDCNRI"
2512..3353
3'UTR          1108 a      560 c      749 g      936 t
BASE COUNT
ORIGIN

Query Match      8.3%;  Score 130;  DB 82;  Length 3353;
Best Local Similarity 100.0%;  Pred. No. 1.11e-71;
Matches 130;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Db 3224 ttggtctcccatcatttaataatgatatattataataataacatcatcattagaaa 3283
Cp 1548 TTGGTTCCTCCATACATTTTAATATGATGATATATTAATCAACATCATCATAGAAA 1489
Db 3284 gcatattacatcatgtttatacaataagcattacattttttaaataaaatgtatacagg 3343
Cp 1488 GCATATTACATACATGTTTATACATAAGCATTACATTTTTTTTAATAAAATCTATACAGG 1429
Db 3344 tggggcaactg 3353
|||||
Cp 1428 TGGGGCACTG 1419

RESULT 4
LOCUS RNDRI42 2206 bp RNA ROD 22-NOV-1996
DEFINITION R.norvegicus mRNA for ER transmembrane protein.
ACCESSION Y07783
NID 91684744
KEYWORDS Dri 42 gene; ER-transmembrane protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2206)
AUTHORS Barila,D.,Plateroti,M., Nobili,F., Muda,A.O., Xie,Y., Morimoto,T.
and Perozzi,G.
TITLE The Dri 42 gene, whose expression is up-regulated during epithelial
differentiation, encodes a novel endoplasmic reticulum resident
transmembrane protein
J. Biol. Chem. 271 (47), 29928-29936 (1996)
MEDLINE 97094703
REFERENCE 2 (bases 1 to 2206)
AUTHORS Perozzi,G
TITLE Direct Submission
Submitted (04-SEP-1996) G. Perozzi, Istituto Nazionale Della
Nutrizione, Unit Of Experimental Nutrition, Via Ardeatina 546,
00178 Roma, ITALY
FEATURES
source
Location/Qualifiers
1..2206
/organism="Rattus norvegicus"
/strain="Wistar"
/cell_type="epithelial"
/dev_stage="adult"
/tissue_type="small intestine"
410..1348
/gene="Dri 42"
/codon_start=1
CDS

```

```
/do_xref="PD:gi/102.6"
/translation="MONEYKDRAIIPESKNGSGSPALNNPRSRSGKRVLLICLDLCLFCE
FMAGLPFLIIETSTIKPIHRGFCYNDESIKPKLTGETINDAVLCAGVIALAIIT
GYPRIYLLKKSSTIONPYAALYKQVCLFGCAISQSFTDIAKVSYLRPLRPHLS
VENVPDQINCSEGYIQNTKRGDSSQVQEARKFSGHASFSTWTLVLYLVLIQARF
```

```

BASE COUNT      327 a      427 c      317 t
TWGACRCGPGCSP*
VCPNDFSCNCSGYIQNTICRGDDSKVOEARCKFFSGHASFMTYLLVLYLQARF
GEYRIYLLKKSSTIONPYVAALYKQVCCFGCAISQFTDIKVSICRLRPHLS
FMAGLPLIETETIKPYHRGVCNDESIKPLKTGETINDAVCAVGIATIALIT
/translation="MONEYKRAIVPESKNGSGPALNNRRSSKRVLLICJDLCEFL
/ad_xref="PD:gi/102/e"

```

GEFYRIYLLKKSRSTIQNPVVAALYKQVCCFLFSCAISQSFTDIAKVSIGRLRPHFLS  
VCNPDFSQINCSEGYIQNYRCRGDDSKVQEARKSFFSGHASFMSYMLVLYLVLYQARF

BASE COUNT	327 a	427 c	373 g	317 t
ORIGIN	TWGARGCSPSCSSP*			
Query Match	6.5%; Score 102; DB 78; Length 1444;			
Best Local Similarity	62.8%; Pred. No. 1.29e-50;			
Matches 263;	Conservative 0; Mismatches 155; Indels 1; Gaps 1;			

Query Match	6.5%	Score 102;	DB 78;	Length 1444;
Best Local Similarity	62.8%	Pred. No. 1.29e-50;		
Matches 263;	Conservative	0;	Mismatches 155;	Indels 1;
				Gaps 1;

Query Match	6.5%;	Score 102;	DB 78;	Length 1444;
Best Local Similarity	62.8%;	Pred. No. 1.29e-50;		
Matches	263;	Mismatches 155;	Indels 1;	Gaps 1;
	Conservative			

Matches	263	Conservative	0	Mismatches	155	Indels	1	Gaps	1
---------	-----	--------------	---	------------	-----	--------	---	------	---

AUTHORS Hunt,T. and Standart,N.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-1990) Hunt T., Standart N., Department of  
Biochemistry, Tennis Court Road, Cambridge, CB2 1QW, England  
REFERENCE 2 (bases 1 to 1727)  
AUTHORS Standart,N., Dale,M., Stewart,E. and Hunt,T.  
JOURNAL Genes Dev. 4 (1990) In press  
COMMENT Forms a complex with the product of the large M1 subunit of  
ribonucleotide reductase.  
FEATURES  
source Location/Qualifiers  
1..1727  
/organism="Spisula solidissima"  
/dev\_stage="oocyte"  
/cell\_type="oocyte"  
96..1250  
/EC\_number="1.17.4.1"  
/codon\_start=1  
/product="M2 (small) subunit of ribonucleotide reductase"  
/db\_xref="PID:gi0342"  
/db\_xref="SWISS-PROT:P07201"  
/translat="M2S11RKESNLGKMKITEENPKKVLGEITNFORSTQK  
TPKQEKIPVKVKSQOVELLADNPRFRVLP IQVHDIWKYKKAESAFTABEVDLS  
KDMAWESLKKKEKH ISHVLAFPAASDGI VNEI VSFKEOVTEARCFYGFQIAM  
AFAHVEGIFSGSFAISFWLKKRGIMPGLTSPESNELI SEDEGLHCDFACLAPSHLVNKV  
ENTRESEVSLIDITYKDQPERDFLNAETMPCVKERADWAMWINDSSSAVERV  
SQERHIOIDEAVKIQVLTETALPCRLIGNCDLMROYISFVADRLLLELKKCDKLYN  
KENPFDNIESUEGKTNFFERKVEYQKMGVNSGNTGDSHAFTLDADF"  
polya\_signal  
BASE COUNT 586 a 260 c 324 g 557 t  
ORIGIN  
Query Match 2.0%; Score 32; DB 45; Length 1727;  
Best Local Similarity 77.6%; Pred. No. 4.10e-03;  
Matches 45; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
  
Db 1670 tgattattaatgaatgtataataatgaatgaatgtgtgnaaaaaaaaaaaaaaaaaa 1727  
||||| ||| ||||| || ||||| ||| ||| ||| ||||| ||||| |||||  
QY 1506 TGATTAAATAATACATATATAAATGTATGGGAGAACCAAAAAAAAAAAAAA 1563  
  
RESULT 9  
LOCUS DMU11584 4601 bp DNA INV 23-JUL-1994  
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.  
ACCESSION U01584  
NID 9508826  
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.  
SOURCE fruit fly.  
ORGANISM Mitochondrion Drosophila melanogaster  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;  
Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 4601)  
Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.  
Sequence, Organization and Evolution of the A+T Region of  
Drosophila melanogaster Mitochondrial DNA  
Mol. Biol. Evol. 11, 523-538 (1994)  
94285822  
2 (bases 1 to 4601)  
Kaguni,L.S.  
Direct Submission  
AUTHORS Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of  
JOURNAL Biochemistry, Michigan State University, East Lansing, MI,  
48824-1318, USA  
FEATURES  
source Location/Qualifiers  
1..4601  
/organism="Drosophila melanogaster"  
/mitochondrion  
/strain="Oregon-R"  
/dev\_stage="embryo"  
550..1022  
/gene="mt:ori"  
/note="Description: mitochondrial origin; repeat I-A"  
/rept\_type=tandem

```

/db_xref="FlyBase:FBgn0013687"
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat I-B1"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
1361..1705
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat I-C/A"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
1706..2043
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat I-B2"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
2044..2388
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat I-C"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
2491..2511
/genes="mt:ori"
/notes="Description: mitochondrial origin; deoxythymidylate stretch"
/db_xref="FlyBase:FBgn0013687"
2512..2648
/partial
/genes="mt:ori"
/notes="Description: mitochondrial origin"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
2649..3112
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat II-A"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
3113..3576
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat II-B1"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
3577..4040
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat II-B2"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
4041..4504
/genes="mt:ori"
/notes="Description: mitochondrial origin; repeat II-C"
/rpt_type=tandem
/db_xref="FlyBase:FBgn0013687"
complement(4565..4585)
/genes="mt:ori"
/notes="Description: mitochondrial origin; deoxythymidylate stretch"
/db_xref="FlyBase:FBgn0013687"
2271 a 131 c 74 g 2125 t

```

Query Match 2.08; Score 31; DB 40; Length 4501;

Best Local Similarity 65.0%; Pred. No. 1.42e-02;

Matches 78; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

Db 2488 ttatttttttttttttttttttttttaataataataataataataataataata 2547

|| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 1560 TTTTITTTTTTTTGGTCTCCCATACATTTTATATGATATATATTAATCAACAT 1501

|| || || || || || || || || || || || || || || || || || || || ||

Db 2548 tatattattgaatttatataatatatatataatagaaaaataattattttataataa 2607

|| || || || || || || || || || || || || || || || || || || || ||

Cp 1500 CAT-TCATAGAAAGCATATTACATACATGTTTATACATTAAGCATTACATTTTAAATAA 1442

```

RESULT 10
LOCUS BMOFLCEG 14626 bp DNA INV 07-MAY-1993
DEFINITION Bombyx mori fibroin light-chain gene, complete cds.
ACCESSION M76430
NID g289362
KEYWORDS fibroin light chain.
SOURCE Bombyx mori 5th instar larvae posterior silk gland DNA.
ORGANISM Bombyx mori
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Insecta; Pterygota; Lepidoptera; Bombycoidea;
Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 14626)
AUTHORS Kikuchi.Y., Mori.K., Suzuki.S., Yamaguchi.K. and Mizuno.S.
TITLE Structure of the Bombyx mori fibroin light-chain-encoding gene:
upstream sequence elements common to the light and heavy chain
Gene 110, 151-158 (1992)
JOURNAL MEDLINE
FEATURES Location/Qualifiers
1..14626
/organism="Bombyx mori"
/dev_stage="5th instar larvae"
/tissue_type="posterior silk gland"
/map="19.2"
900..907
CAAT_signal
997..1003
TATA_signal
1028..1104
exon
/genes="Fib-L"
/number=1
join(1028..1104,9250..9332,10532..10727,11412..11555,
12196..12284,13361..13505,14061..14499)
/genes="Fib-L"
/product="fibroin light-chain"
join(1069..1104,9250..9332,10532..10727,11412..11555,
12196..12284,13361..13505,14061..14156)
/genes="Fib-L"
/codon_start=1
/product="fibroin light-chain"
/db_xref="PID:g289363"
/translaton="MKPIFLVLVATSAAPSVTINQYSDNEIPRIDDGKASSVIS
RWKDYVDITDKSIATLNVQELKDMASQGYASQASAVATAGIIAHLASGPGDACA
AAVNVNSTDGVRSNCFAGRSILGPFEGHVGONLNLINOLVNPGLVSYVGPALGC
AGGRFYDEFAANDAILASSDSSFLNEEYCIKVELVNSRSQSNNTAAITAHLLPPV
AQVFHQSGSITDLLRGVGNQNDATGLVANQRYIAQNASQVHV"
1105..9249
/genes="Fib-L"
/number=1
9250..9332
/genes="Fib-L"
/number=2
9333..10531
/genes="Fib-L"
/number=2
10532..10727
/genes="Fib-L"
/number=3
10728..11411
/genes="Fib-L"
/number=3
11412..11555
/genes="Fib-L"
/number=4
11556..12195
/genes="Fib-L"
/number=4
12196..12284
/genes="Fib-L"
/number=5
12285..13360
/genes="Fib-L"
/number=5
13361..13505
/genes="Fib-L"
/number=6

```



intron 13506..14060  
 /gene="Fib-L"  
 /number=6  
 exon 14061..14499  
 /gene="Fib-L"  
 /number=7  
 polyA\_signal 14473..14478  
 /gene="Fib-L"  
 14499  
 polyA\_site /gene="Fib-L"  
 4691 a 2726 c 2643 g 4566 t  
 BASE COUNT  
 ORIGIN  
 Query Match 2.0%; Score 31; DB 35; Length 14626;  
 Best Local Similarity 70.0%; Pred. No. 1.42e-02;  
 Matches 63; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
 Db 3136 atacgtttattatttttaataattttataattattattt-gaaattattttatata 3194  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1536 ATACGTTTAAATGATTATTTATTTAAATCAACATCATTCATAGAAGCATATTACATA 1477  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 3195 catattataaaatagtttttaataattatt 3224  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Cp 1476 CATGTTTATACATAGCATTCATACATTTT 1447  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 RESULT 11  
 LOCUS DMU37541 19517 bp DNA INV 28-FEB-1996  
 DEFINITION Drosophila melanogaster mitochondrial genome, complete sequence.  
 ACCESSION U37541  
 MID g1166529  
 KEYWORDS A+T-rich region; ATPase; circular; complete genome; cytochrome b;  
 cytochrome c; cytochrome c oxidase; mitochondrial DNA; NADH  
 dehydrogenase; ribosomal RNA; transfer RNA-Ala; transfer RNA-Arg;  
 transfer RNA-Asn; transfer RNA-Asp; transfer RNA-Cys; transfer  
 RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His;  
 transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer  
 RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser;  
 transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer  
 RNA-Val.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Mitochondrion Drosophila melanogaster  
 Eukaryotae; Mitochondrial eukaryotes; Metazoa; Arthropoda;  
 Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;  
 Ephyroides; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 408; 13319 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Drosophila melanogaster mitochondrial DNA: completion of the  
 nucleotide sequence and evolutionary comparisons  
 JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)  
 MEDLINE 96423163  
 REFERENCE 2 (bases 404 to 5272)  
 AUTHORS de Bruijn,M.H.  
 TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and  
 genetic code  
 JOURNAL Nature 304 (5923), 234-241 (1983)  
 MEDLINE 83245048  
 REFERENCE 3 (bases 5269 to 5695)  
 AUTHORS Clary,D.O., Wahlethner,J.A. and Wolstenholme,D.R.  
 TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5'  
 flanking sequences and comparisons to mammalian mitochondrial tRNA  
 genes  
 JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)  
 MEDLINE 83220794  
 REFERENCE 4 (bases 12511 to 12682)  
 AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and  
 Wolstenholme,D.R.  
 TITLE Drosophila mitochondrial DNA: a novel gene order  
 JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)  
 MEDLINE 83090428  
 REFERENCE 5 (bases 804 to 1778)  
 AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.  
 TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in in

Drosophila melanogaster and its sibling species  
 Mol. Biol. Evol. 4 (6), 638-650 (1987)  
 MEDLINE 88174373  
 REFERENCE 6 (bases 441 to 2967)  
 AUTHORS Satta,Y. and Takahata,N.  
 TITLE Evolution of Drosophila mitochondrial DNA and the history of the  
 melanogaster subgroup  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)  
 MEDLINE 91088557  
 REFERENCE 7 (bases 5268 to 13619)  
 AUTHORS Garesse,R.  
 TITLE Drosophila melanogaster mitochondrial DNA: gene organization and  
 evolutionary considerations  
 JOURNAL Genetics 118 (4), 649-663 (1988)  
 MEDLINE 88212147  
 REFERENCE 8 (bases 14215 to 14512)  
 AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and  
 Atkinson,P.W.  
 TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are  
 modified arthropods  
 JOURNAL Science 258 (5086), 1345-1348 (1992)  
 MEDLINE 93088057  
 REFERENCE 9 (bases 14917 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.  
 TITLE Sequence, organization, and evolution of the A+T region of  
 Drosophila melanogaster mitochondrial DNA  
 JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)  
 MEDLINE 94285822  
 REFERENCE 10 (bases 1 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,  
 Michigan State University, East Lansing, MI 48824-1319, USA  
 FEATURES  
 source  
 1..19517  
 /organism="Drosophila melanogaster"  
 /mitochondrion  
 /note="derived from new and previously submitted  
 sequences; sequence is a composite containing sequences  
 obtained from different Drosophila melanogaster strains"  
 1..65  
 /gene="mt:ND6"  
 /note="Description: mitochondrial NADH-ubiquinone  
 oxidoreductase chain 6"  
 /db\_xref="FlyBase:FBgn0013685"  
 /product="tRNA-Ile (CAU)"  
 complement(97..165)  
 /gene="mt:ND6"  
 /note="Description: mitochondrial NADH-ubiquinone  
 oxidoreductase chain 6"  
 /db\_xref="FlyBase:FBgn0013685"  
 /product="tRNA-Gln (UUG)"  
 171..239  
 /gene="mt:ND6"  
 /note="Description: mitochondrial NADH-ubiquinone  
 oxidoreductase chain 6"  
 /db\_xref="FlyBase:FBgn0013685"  
 /product="tRNA-f-met (CAU)"  
 240..1265  
 /gene="mt:ND6"  
 /note="Description: mitochondrial NADH-ubiquinone  
 oxidoreductase chain 6"  
 /codon\_start=1  
 /db\_xref="NADH dehydrogenase subunit 2"  
 /product="NADH dehydrogenase subunit 2"  
 /transl\_table=5  
 /transl\_table=5  
 /translation="MFNNSKILFTIMIGTLITVTSWIGAMWGIEINLLSPIL  
 LSDNNIMSTASLKFTQVLASTVLLSSILLKNNNNNEINESTSMISALL  
 LKSGAAPFFWFPNMEGLTWMALMTWQKIAPMLISYINIKYLLISVILSVII  
 GAIGGLAFTSRLKLMATSSINHLGMLSLMISEIWLIFFFYFLGFLVTFMENIF  
 KLFHNLQFSWFVNSKILKTFPMNTLSGLGLPPFLPKVLVQQLTCNQYFMLT  
 IMNSTLITLFFYLRCYSAFMNYFENNWMKMNNSINYNMYIMIFFSIFGLFI



[illegible]

REFERENCE  
2 (bases 1 to 1862)  
/RO, UK  
AUTHORS  
Millward-Sadler, S. J., Hall, J., Black, G. W., Hazlewood, G. P. and



\*\*\*\*\*  
MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 10:11:33 1997; MasPar time 169.52 Seconds  
Tabular output not generated. 962.304 Million cell updates/sec  
Title: >US-08-842-827-1  
Description: (1-1563) from US08842827.seq  
Perfect Score: 1563  
N.A. Sequence: 1 CCTGTGGGAGAGACGCGCG .....CCAAAAA.....  
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTTTTT  
Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 142080 seqs, 52183452 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: n-genesq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29  
Statistics: Mean 9.613; Variance 6.601; scale 1.456  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

c	16	34	2.2	114	12	Q70469	Generic DNA sequence	2.07e-04
	17	33	2.1	198	7	Q42784	Ligand-induced gene,	6.07e-04
	18	32	2.0	114	12	Q70466	Generic DNA sequence	1.77e-03
c	19	31	2.0	114	12	Q70470	Generic DNA sequence	5.07e-03
	20	32	2.0	114	12	Q70466	Generic DNA sequence	1.77e-03
	21	31	2.0	3521	7	Q41226	Clone GP3 encoding ma	5.07e-03
	22	29	1.9	114	12	Q70472	Generic DNA sequence	4.02e-02
c	23	30	1.9	114	12	Q70472	Generic DNA sequence	1.44e-02
c	24	29	1.9	114	12	Q70473	Generic DNA sequence	4.02e-02
c	25	29	1.9	1337	23	T33852	Rat vas deferens P2x	4.02e-02
	26	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
c	27	28	1.8	114	12	Q70471	Generic DNA sequence	1.11e-01
c	28	28	1.8	565	6	Q35072	HCV envelope region n	1.11e-01
	29	28	1.8	1174	9	Q51020	Human FcγRII alpha gen	1.11e-01
c	30	26	1.7	74	21	T13613	DC43 TSAR library gen	8.06e-01
c	31	26	1.7	81	21	T13611	DC43 TSAR library gen	8.06e-01
	32	27	1.7	347	4	Q27482	T23 cDNA.	3.02e-01
	33	27	1.7	498	3	N50034	Sequence encoding new	3.02e-01
	34	27	1.7	501	3	N50024	Sequence encoding new	3.02e-01
c	35	27	1.7	501	3	N50023	Sequence encoding new	3.02e-01
	36	26	1.7	501	3	N50033	Sequence encoding new	8.06e-01
	37	27	1.7	3520	11	Q69215	Interleukin-10 recept	3.02e-01
	38	26	1.7	4193	14	Q82834	Human stromalin-2 DNA	8.06e-01
	39	25	1.6	39	7	Q51787	Mixed oligonucleotide	2.12e+00
	40	25	1.6	501	3	N50031	Sequence encoding new	2.12e+00
	41	25	1.6	501	3	N50026	Sequence encoding new	2.12e+00
	42	25	1.6	501	3	N50025	Sequence encoding new	2.12e+00
	43	25	1.6	587	9	Q55685	Rabbit CAP18 gene.	2.12e+00
	44	25	1.6	961	2	Q10791	Bovine vascular endot	2.12e+00
	45	25	1.6	1633	17	T15608	FGF receptor ligand F	2.12e+00

ALIGNMENTS

RESULT 1  
ID Q10572 standard; DNA: 1047 BP.  
AC Q10572;  
DT 09-APR-1991 (first entry)  
DE Human Natriuretic Peptide Receptor B.  
KW NPRB; ANP; BNP; kidney failure; heart failure; protein kinase;  
OS hyperaldosteronism; glaucoma; guanyl cyclase.  
FH Homo sapiens.  
FT Key: Location/Qualifiers  
FT Peptide 1..22  
FT /label= signal sequence  
FT Protein 12  
FT /label= mature NPBR  
FT Domain 23..455  
FT /label= extracellular domain  
FT /note= "binds natriuretic peptides A,B and C]"  
FT Domain 456..456  
FT /label= transmembrane domain  
FT Domain 479..1047  
FT /label= cytoplasmic domain  
FT /note= "GC and protien kinase activity"  
FT Modified -site 24..26  
FT /label= N-glycos\_site  
FT Modified -site 35..37  
FT /label= N-glycos\_site  
FT Modified -site 161..163  
FT /label= N-glycos\_site  
FT Modified -site 195..197  
FT /label= N-glycos\_site  
FT Modified -site 244..246  
FT /label= N-glycos\_site  
FT Modified -site 277..279  
FT /label= N-glycos\_site  
FT Modified -site 349..351  
FT /label= N-glycos\_site  
FT Modified -site 600..602  
FT /label= N-glycos\_site  
PN WO9100292-A.  
PD 10-JAN-1991.

```

PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB: Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 4.9%; Score 77; DB 2; Length 1047;
Best Local Similarity 9.0%; Pred. No. 4.56e-27;
Matches 73; Conservative 213; Mismatches 517; Indels 5; Gaps 5;

Db 170 yndatrdhnytnngvnanagonsvnhvnyarngngnnnathnrangrnynacgn 229
Cp 1179 TCGGATAGTATCCCGATGTTGGTGTTCATGCAGAGTGTATGAGAGTCCTCTTT 1120
Db 230 nmhnnnnnnaarntngdvnyndvngnsragntrtrtgrndvndtrtnnanarna 289
Cp 1119 TTTCTTTCTTTAAAGAAAGTTCTTTCTTTGAAGAAATCCGATACATACAGCAATAA 1060
Db 290 nntvntvnyrnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 349
Cp 1059 TTGCAACACAGAGTCCTCGATGATGATGATGATGATGATGATGATGATGATGAT 1000
Db 350 nttnnggtrndgrnrvnkmgrryghvgtgnvmdkndrntdnvngwngdndsgdnnaa 409
Cp 999 CAGAAACTCGAAGAGCCCAATGATGATGATGATGATGATGATGATGATGATGATG 940
Db 410 hyssganknnwtgrnnnnwvgnnsdnnncandndndndndndndndndndndnd 469
Cp 939 GCGGTA-AGAGTCTGCGGAGTCTCCCTTCATCTCG-GGTTGAAGATAAGTGCCACAA 882
Db 470 ngvsennnnrknmknknnaamwrrnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 529
Cp 881 CAGCATGAGTACATGGAAGAAACGAGAGTGGCTGATGATGATGATGATGATGATGAT 822
Db 530 ahgkyunnantghnkvavankhvkkrntrnnnnnnnnnnnnnnnnnnnnnnnnnn 589
Cp 821 AACTCTTTTCGATTCCTCGACATATGATGATGATGATGATGATGATGATGATGATG 762
Db 590 nnnncvntvcyngnsndnnnd-snnndwmrnyssndndvknmnnnnnnnnnnnnnn 648
Cp 761 TTTTGACCAATCTGGATCAACAACATCAAGAAGTGGCCGAGTCTGCTGCTGCTGCT 702
Db 649 ncvdsrvnkvntdygnasrnsrtnndnnnnyaknnntennnnnnnnnnnnnnnn 708
Cp 701 CTTGGCATGTGCTAGGAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGTACT 643
Db 709 sngnnnnnnnrgnyngndnsnknvkvngnrgnryrnsndrtnnnnnnnnnnnnnnn 768
Cp 642 CTTTCTAAATAGTGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 583
Db 769 wandnanrndngnnknrrnnknnggtsndnnnnnnnnnnnnnnnnnnnnnnnnnn 828
Cp 582 AGTAACACAGACAGGGTTCTCCAGATAATATACGATATATCTGATGATGATGATGAT 523
Db 829 kanannnnnnhsvannknrgntvnaandsvtnynsdhvgntansntmmvntnnd 888
Cp 522 CACCTTA-ATAACGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 464
Db 889 nytcdannndndvykvntngdaymvvsgnngnrgnrrhannnnnnnnnnnnnnnn 948

```

```

Cp 453 TTACAGATATCTCTCTGTTGGAGGGGTATCCCTTGAAGTAAGATTTGCAAAAGGCAAT 404
Db 949 rnhdnrrnngvhtgnvcagvvgvgnknmr 976
Cp 403 CCAGCCAGCAACACGAGCAGACATCGA 376

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanylyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protein kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 600..602
FT /label= N-glycos_site
FT WO9100292-A.
PN 10-JAN-1991.
PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB: Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 4.7%; Score 74; DB 2; Length 1047;
Best Local Similarity 7.9%; Pred. No. 2.16e-25;
Matches 60; Conservative 215; Mismatches 479; Indels 5; Gaps 5;

Db 278 rtrnnanannntvntvnyrnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnsgcn 337

```



01-DEC-1993.  
PF 24-MAY-1993; 108325.  
PR 26-MAY-1992; US-889651.  
PA (BECT ) BECTON DICKINSON CO.  
PI Shank DD, Spears PA;  
WP1: 93-378844/48.  
PT New oligo:nucleotide probes specific for Mycobacteria - used for  
detection and amplification of Mycobacteria nucleic acid in  
samples  
PS Claim 3; Page 14; 23pp; English.  
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
(Q51735). It hybridized to all spp. of mycobacteria tested, but  
cross reacted to a few non-mycobacterial spp. The probe may  
be useful as an initial screen for mycobacterial infection.  
CC See also Q51735-45 and Q51747-59.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.6%; Score 41; DB 9; Length 91;  
Best Local Similarity 10.0%; Pred. No. 8.33e-08;  
Matches 6; Conservative 47; Mismatches 6; Indels 1; Gaps 1

Db 7 ggcgssvhsyvv-vhvsvshhsbhvhvhhvsvvvvhhvhhvhhvhhvhyvsvctcaa 65  
|| :  
QY 1124 GGAGGACTCTCATCACTCGATGGAACACCACAACACTGGGATCCTATTCGAGCAA 1183

RESULT 6  
ID N81164 standard; DNA; 204 BP.  
AC N81164;  
DT 08-NOV-1990 (first entry)  
DE Base substituted E.coli beta-galactosidase alpha-fragment.  
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.  
OS Escherichia coli.  
FH Key Location/Qualifiers  
FT misc\_feature 19..69  
FT /\*tag\_a  
FT /function=multipe cloning site  
FT primer\_bind 187..204  
FT /\*tag\_b  
PN EP-285123-A.  
PD 05-MAR-1988;  
PF 30-MAR-1988; 105163.  
PR 03-APR-1987; US-034819.  
PA (SUSO) SUOMEN SOKERI OY.  
PI Lehtoavara P, Knowles J, Koivula A, Bamford J, Reinikainen T;  
WI: 88-279927/40.  
PT Introducing random point mutations into nucleic acids -  
by prepn of single stranded template, annealing a primer, elongation,  
misincorporation, completion of molecules and screening.  
PS Disclosure; p: English.  
CC Random point mutations were introduced into the alpha fragment of  
E.coli beta-galactosidase. The wild type sequence was obtained as a  
single stranded template and an oligonucleotide was hybridised to  
it to generate a popn of DNA molecules which terminate at all  
possible nucleotide positions within a specified region. The  
variable 3' ends generated in this way are used as primers for  
reverse transcriptase. Nucleotides are misincorporated by the  
transcriptase and the molecules are completed to forms that can be  
amplified and then expressed in a suitable host-vector system.  
CC The sequence covers all 176 diff base substitutions, most of which  
occurred singularly in any given mutant.  
CC See also P80575.  
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.4%; Score 38; DB 1; Length 204;  
Best Local Similarity 19.5%; Pred. No. 2.50e-06;  
Matches 31; Conservative 63; Mismatches 64; Indels 1; Gaps 1

Db 34 bgcagcgacbcyrragnnycccggygccgcycaayyddchvcgcymrttthh 93  
||||| ||||| :  
Cc 300 TGGCGACGACGCCGCAGCG-CCCATTGATTATCGAGCCCCGCGGGGTGCGAGGCC 242

Db 94 yrrmrbnvyrdynrdaawyyccrrsvkydcvcynachdddhvbbvbnvnmhncnc 153





PT screening a recombinant vector library expressing fusion proteins  
PT comprising a binding domain and an effector domain  
PS Disclosure: Page 35; 255pp; English.  
PS 070469 is a generic DNA sequence used to generate random TSAR peptide  
CC This generic formula can be represented as follows: X(TGC)(NNB)10-  
CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction  
CC sites (X is not the same as Y) that are not specified further. This  
CC sequence generates peptides that are cloverleaf in structure. Other  
CC generic sequences are shown in 070465-68. Other specific peptides  
CC generated by these generic sequences are shown in R65150-54. TSARs are  
CC concatenated heterofunctional proteins or peptides, comprising at least  
CC two functional regions - a binding domain with affinity for a ligand and  
CC a second effector peptide portion that is chemically or biologically  
CC active. They may further comprise a linker peptide between the 2 domains.  
CC The oligonucleotides are also designed so that the expressed peptide  
CC contains 2 or 4 cysteine residues positioned in, or flanking, the  
CC unpredicted or variant residues. These residues confer some degree of  
CC conformational rigidity to the peptides. The TSARs or compens. comprising  
CC a TSAR binding domain can be used in vivo to deliver a chemically or  
CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin  
CC or enzyme, to the specific target or on the cell. They can also replace  
CC the function of macromolecules, eg. monoclonal or polyclonal antibodies  
CC and therefore circumvent the need for complex methods of hybridoma  
CC formation or in vivo antibody production. The TSARs are easily  
CC characterised and have designed activity allowing direct and rapid  
CC detection in a screening process.  
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

```
Query Match      2.28; Score 35; DB 12; Length 114;
Best Local Similarity 7.18; Pred. No. 6.95e-05;
Matches 8; Conservative 31; Mismatches 74; Indels 0; Gaps 0;
```

```

RESULT      11
ID          Q70465 standard; DNA; 114 BP.
AC          Q70465;
DE          05-APR-1995 (first entry)
DE          DE Generic DNA sequence to generate a random TSAR peptide library.
DE          TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW          effector domain; concatenated heterofunctional protein; linker;
KW          direct; rapid; detection; screening; treatment; generic; ss.
OS          Synthetic.
FH          Key
FH          Location/Qualifiers
FT          misc_feature      55..60
FT          /tag= a
FT          /note= "this sequence represents 'Z'; Z can be a
FT          sequence of 6, 9 or 12 nucleotides (see
FT          comments)"
FT          PN          WO9418318-A.
PD          18-AUG-1994.
PF          01-FEB-1994;      U00977.
PR          01-FEB-1993;      US-013416.
PR          30-DEC-1993;      US-176500.
PR          31-JAN-1994;      US-189331.
PA          (UYNC-) UNIV NORTH CAROLINA.
PA          Fowlkes DM, Kay BK;
PI          WPI: 94-279739/34.
DR          P-PSDB; R65150 and R65151.
PT          Identifying proteins or peptide(s) which bind a ligand - by
PT          screening a recombinant vector library expressing fusion proteins
PT          comprising a binding domain and an effector domain
PT          Disclosure; Page 35; 253pp; English.
CC          Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC          Synthetic Affinity Reagents) peptides.This generic formula can also be
CC          represented as follows: X(NNB)6(TGC)(NNB)112((NNB)14(TGC)(NNB)3Y. X

```

CC and Y are flanking restriction sites (X is not the same as Y) that are not specified further. Other generic sequences are shown in Q70466-68. Other specific peptides generated by these generic sequences are shown in R51511-54. TSARs are concatenated heterofunctional proteins or peptides, comprising at least two functional regions - a binding domain with affinity for a ligand and a second effector peptide portion that is chemically or biologically active. They may further comprise a linker peptide between the 2 domains. The oligonucleotides are also designed so that the expressed peptide contains 2 or 4 cysteine residues positioned in, or flanking, the unpredicted or variant residues. These residues confer some degree of conformational rigidity to the peptides. The TSARs or compans, comprising a TSAR binding domain can be used in vivo to deliver a chemically or biologically active moiety, eg. metal ion, radioisotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg. monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or in vivo antibody production. The TSARs are easily characterised and have designed activity allowing direct and rapid detection in a screening process.

SQ Sequence 114 BP; 2 G; 2 C; 2 G; 2 T;

Query Match 2.2%; Score 34; DB 12; Length 114;  
Best Local Similarity 3.6%; Pred. No. 2,078-04;  
Matches 4; Conservative 33; Mismatches 75; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62  
QY 106 GCCTCGGCATTAACCGAGTTCGCGGGGGCTGTGAGGGAGGCCCGCGCATTCG 165  
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114  
QY 166 TGGCGGTGGAGGCGCGCGGTCTCAGCGCGCCTCGGCTGCTCTCTCTCT 217

RESULT 12

ID	Q70467 standard; DNA; 114 BP.
AD	Q70467;
DE	05-APR-1995 (first entry)
DB	Generic DNA sequence to generate a random TSAR peptide library.
DT	TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW	effector domain; concatenated heterofunctional protein; linker;
KW	direct; rapid; detection; screening; treatment; generic; ss.
OS	Synthetic.
FH	Key
FT	misc.feature 55..60
FT	/tag= a
FT	/note= "this sequence represents 'Z'; Z can be a
FT	sequence of 6, 9 or 12 nucleotides (see
FT	comments)"
PN	WQ9418318-A.
PD	18-AUG-1994.
PF	01-FEB-1994; U00977.
PR	01-FEB-1993; US-013416.
PR	30-DEC-1993; US-176500.
PR	31-JAN-1994; US-189331.
PA	(UWNC) UNIV NORTH CAROLINA.
PI	Fowlkes DM, Kay BK;
DR	WPI: 94-279739/34.
DR	P-PSDB; R65153.
PT	identifying proteins or peptide(s) which bind a ligand - by
PT	screening a recombinant vector library expressing fusion proteins
PT	comprising a binding domain and an effector domain
PS	Disclosure; Page 35; 255pp; English.
CC	Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC	Synthetic Affinity Reagents) peptides. This generic formula can also be
CC	represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TCC)(NNB)1Y X
CC	and Y are flanking restriction sites (X is not the same as Y) that are
CC	not specified further. Other generic sequences are shown in Q70466-68.
CC	Other specific peptides generated by these generic sequences are shown in
CC	R51511-54. TSARs are concatenated heterofunctional proteins or peptides,
CC	comprising at least two functional regions - a binding domain with
CC	affinity for a ligand and a second effector peptide portion that is
CC	chemically or biologically active. They may further comprise a linker



Search completed: Thu Nov 6 10:14:51 1997  
Job time : 198 secs.

Cp 150 GCCTCCCTCACAGCCCGCGAACAACCTGGTTAGTGCGGAGCGCTGTTGCCAGCG 91  
 Db 63 bnnbnnbnnbnnbcaacnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbcaacacnnb 114  
 Cp 90 GCGCGAGCTCTAGCTCAGGACCTCTCAGCGCGGACGCGCTGCCCGGT 39

```

RESULT 15
ID Q70468 standard; DNA: 114 BP.
AC Q70468;
DE OS-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR petide library.
DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key
FH Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /*note= "this sequence represents 'Z'; Z can be a
FT : sequence of 6, 9 or 12 nucleotides (see
FT comments)"
FN W09418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UNIC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
PI WPI: 94-279739/34.
PI P-PSDB; R65154.
DR Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 253pp; English.
CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides.This generic formula can also be
CC represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC Q8511-34. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector portion that is
CC chemically or biologically active.They may further comprise a linker
CC peptide between the 2 domains.The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compans. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have detected activity
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

```

	Query Watch	2.28;	Score 34;	DB 12;	Length 114;	
	Best Local Similarity	4.5%;	Pred. No. 2.07e-04;			
	Matches	5;	Conservative	32;	Mismatches 75;	Indels 0; Gaps 0;
D <b>b</b>	3 bnnbnbnbnbnbnbnbnbnbnbnbnbnbtgcnnbnbnbnbnbnbnbnbnbnbnbnbn					
C <p></p>	177 CTCCCAACGCCAGCAATGGCGCGCGGGGCGCTCCCTCAGACGCCGCCGGAACACTCGGT	118				
D <b>b</b>	63 bnnbnbnbnbnbnbnbnbnbtgcnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn	114				
C <p></p>	117 TAGTGCAGGAGCGCTGCTGTGTCAGACGCCGCGCACTCTCTAGCTTCAGGACC	56				

\*\*\*\*\*  
M P S R E H  
\*\*\*\*\*  
(TW)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 10:32:01 1997; MacPar time 427.87 Seconds  
Tabular output not generated.  
1003.878 Million cell updates/sec

Title: >US-08-842-827-1  
Description: (1-1563) from US08842827.seq  
Perfect Score: 1563  
N.A. Sequence: 1 CCTGTGGGAGAGACGGCCGG.....CCMAAATAAATAAATAA 1563  
Comp: GGACACCTCTCTCGCCGCC.....GGTTTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13

Database: EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnSTS 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnSTS 107:ueEST1  
108:ueEST2

Statistics: Mean 11.832; Variance 3.133; scale 3.777

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.	
1	482	30.8	514	28	AA024599	ze78b09.r1 Soares fet	0.00e+00	
c	380	24.3	390	9	AA132640	zo20d08.r1 Stratagene	0.00e+00	
2	376	24.1	415	87	HS605336	zb63f08.r1 Soares fet	0.00e+00	
3	376	23.0	376	9	AA132534	zo20d08.s1 Stratagene	0.00e+00	
c	359	23.0	376	9	AA132534	zo20d08.s1 Stratagene	0.00e+00	
4	352	22.5	378	28	AA024600	ze78b09.s1 Soares fet	0.00e+00	
c	345	22.1	377	30	AA062861	zf70e10.s1 Soares pin	0.00e+00	
6	286	18.3	312	30	AA062853	zf70d09.r1 Soares pin	0.00e+00	
8	259	16.6	385	33	AA107326	mi95b10.r1 Stratagene	0.00e+00	
c	229	14.7	542	22	AA210789	zr90e08.r1 Soares NBH	4.23e-286	
c	229	14.7	542	54	AA210789	zr90e08.r1 Soares NBH	4.23e-286	
11	123	7.9	411	50	AA152779	mr78c03.r1 Stratagene	5.49e-134	
12	120	7.7	241	87	HS368337	zc68b11.s1 Soares fet	1.56e-129	
13	119	7.6	501	43	G23282	human STS WI-15758	4.73e-128	
c	14	88	313	91	MM155742	mx78g01.r1 Soares mou	6.59e-83	
c	15	88	313	66	AA237956	mx78g01.r1 Soares mou	6.59e-83	
c	16	76	4.9	564	1	W30942	zc64f09.r1 Soares fet	5.40e-66
c	17	76	4.9	564	88	HS942332	zc64f09.r1 Soares fet	5.40e-66
18	58	3.7	87	33	AA106892	mi85a07.r1 Stratagene	1.43e-41	
19	53	3.4	242	27	W39815	308 Mouse VM CDNA lib	4.55e-35	
20	53	3.4	337	27	W39816	309 Mouse VM CDNA lib	4.55e-35	
21	53	3.4	402	27	W39811	307 Mouse VM CDNA lib	4.55e-35	
c	22	46	320	10	AA015411	mh21g01.r1 Soares mou	2.96e-26	
23	44	2.8	478	26	N75714	yy31b01.r1 Soares fet	8.23e-24	
24	42	2.7	185	90	MM0272	ma69e06.r1 Soares mou	2.09e-21	
c	25	40	2.6	184	95	MM63211	mh78e01.r1 Soares mou	4.83e-19
c	26	40	2.6	199	10	AA014140	mh29b02.r1 Soares mou	4.83e-19
c	27	40	2.6	269	95	MM5326	mb22b07.r1 Soares mou	4.83e-19
c	28	37	2.4	52	AA122946	mq69e06.r1 Stratagene	1.38e-15	
c	30	34	2.2	369	24	AA217401	mu90f02.r1 Soares mou	2.98e-12
c	31	34	2.2	429	56	AA144221	mr14h12.r1 Soares mou	2.98e-12
32	33	2.1	236	81	HS1146241	zp52f10.r1 Stratagene	3.59e-11	
33	33	2.1	236	12	AA182434	zp52f10.r1 Stratagene	3.59e-11	
34	33	2.1	301	56	AA138853	mr03a06.r1 Soares mou	3.59e-11	
35	29	1.9	181	67	AA241676	MEL2SJB85T3 JHU96SL-B	4.97e-07	
36	29	1.9	181	79	BMAA1676	MEL2SJB85T3 JHU96SL-B	4.97e-07	
37	29	1.9	197	70	AA208192	mv91b08.r1 GuayWoodfo	4.97e-07	
38	29	1.9	197	103	MMX81921	mv91b08.r1 GuayWoodfo	4.97e-07	
c	40	2.9	339	1	N97773	1238C3 czapPFD2.1, D	4.97e-07	
c	41	30	1.9	446	12	AA181403	zp52f10.s1 Stratagene	4.90e-08
42	28	1.8	153	15	AA189532	mi90c05.r1 Soares mou	4.80e-06	
43	28	1.8	244	5	AA117061	mn29d08.r1 Beddington	4.80e-06	
c	44	28	1.8	453	105	MMW9541	mf84g07.r1 Soares mou	4.80e-06
45	28	1.8	1097	73	C22771	Dictyostelium discoid	4.80e-06	

ALIGNMENTS

1  
RESULT AA024599 514 bp mRNA EST 01-FEB-1997  
LOCUS ze78b09.r1 Soares fetal heart NDBH19W Homo sapiens CDNA clone  
DEFINITION 365081 5'  
ACCESSION AA024599  
NID gl489504  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 514)  
AUTHORS Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,  
Holman, M., Holtman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, B., Ronning, F., Tan, F., Trevisan, E.,  
Waterson, R., Williamson, A., Wohldmann, P. and Wilson, R.  
TITLE WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 600 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 476.

# FEATURES

Location/Qualifiers

1..514

/organism="Homo sapiens"

/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTCAGATGGGAGCGGCCGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot - 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBL19W."

/clone="365081"

/clone\_lib="Soares fetal heart NBLH19W"

/sex="unknown"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

<1..>514

BASE COUNT 162 a 98 c 97 g 153 t 4 others

# FEATURES

source

1..390

/organism="Homo sapiens"

/note="Organ: colon; Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dt. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'." /clone="587439"

/clone\_lib="Stratagene colon (#937204)"

/lab\_host="SOLR cells (kanamycin resistant)"

<1..>390

BASE COUNT 111 a 81 c 82 g 115 t 1 others

ORIGIN

Query Match

Best Local Similarity 24.38; Score 380; DB 9; Length 390;

Matches 387; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 gaaggctggagtttttaataagtttagagcttattagataaccactgagtttaagga 60

Cp 1410 GAAGGCTGGAGTTTTTTTAAAGTTTAGAGCTATTAGATAACCACCTGAGTTAAAGTA 1351

Db 61 actatgtacacacaaagtgtgcaatccaaagggccatagacacagcagcagcttaagg 120

Cp 1350 ACTATGTACACAAAGTGTGCA-TCCAAAGGCGCATATACAGCAGAGCTTTTAAAGG 1292

Db 121 ctgtacacacaggaagaagatgcatcctcttgccttggcatttcttccttttagaa 180

Cp 1291 CTGTACACACGGAAGAAAGATGCATCTCTTGCCTTGTGGCAATCATTTTCTTTAGAA 1232

Db 181 aacaggccagcttcacctggcaccctgctgcttcaagctggtgctgctgtag 240

Cp 1231 AACAGGCGACCTTCACCTGGGCGACCTGCTGCTTTCAAGGCTGGTGAITCTCGGATAG 1172

Db 241 tgattccagttgttggtgttcaatgcagagttgtatgagagtcctcctcttcttct 300

Cp 1171 TCATTCCAGTTGTGGTGTTCATGCAGAGTTGTATGAGAGTCTCTCTCTTTCTTTCT 1112

Db 301 ttaaaagaagttcttcttgaagaataccgatacatatagacaactaatattgcaacc 360

Cp 1111 TTAAGAAGTTCCTTTTGAAGAATAATCCGATACATATACAGCAACTATATATTCGAACC 1052











LOCUS AA210789 542 bp mRNA EST 14-FEB-1997  
 DEFINITION zr90e08.r1 Soares NBHTGBC Homo sapiens cDNA clone 682982 5'.  
 ACCESSION AA210789  
 NID G1809443  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 542)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, B., Rolfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilton RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 414.  
 FEATURES  
 source  
 1..542  
 /organism="Homo sapiens"  
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from germinal B-cells (flow-sorted from tonsils) provided by Dr. Louis Staudt of the NCI, and was then primed with a Not I oligo(dT) primer [5', TGTTCCCAATCGAATGGAGCGGCGCCATTTTTTTTTTTTTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 /clone="682982"  
 /clone\_lib="Soares NBHTGBC"  
 /tissue\_type="Germinal B-cell"  
 /lab\_host="DH10B"  
 <1..>542  
 BASE COUNT 162 a 97 c 117 g 164 t 2 others  
 ORIGIN  
 mRNA  
 14.7%; Score 229; DB 54; Length 542;  
 Best Local Similarity 95.4%; Pred. No. 4.23e-296;  
 Matches 308; Conservative 0; Mismatches 2; Indels 13; Gaps 13;  
 Db 220 ttgtgttcccatatatttaattgtattattataataacaaacattcattatgaaa 279  
 |||||  
 Cp 1548 TTTGGTCTCCCATACATTTTAAATGTATTATATTATTAATCAACATCATTCATAGAA 1489  
 |||||  
 Db 280 gcatattacatcatgtttattacatacaacattacatttttttaataaaaatgtatcacagg 339  
 |||||  
 Cp 1488 GCATATTACATACATGTTTATACATAAGCATATACATTTTATATAAAATGTATACAGG 1429  
 |||||  
 Db 340 tggggcactgttttggcgaaggcttgaggtttttttaattagcttagactattagat 399  
 |||||  
 Cp 1428 TGGGGCACTGTTTGGTGAAGCTTGGAGTTTGTTTT-AATGAGTTTATAGCTATTAGAT 1370  
 |||||  
 Db 400 aaccactgattaaagggttaactatggttacacacaaagggtgtgcattccaaagaagcata 459  
 |||||  
 Cp 1369 AACCTCTGAGTTAAAGG-TAACTATG-TACACAAAG-TGTGC-ATCCAAGA-GGCATA 1315  
 |||||  
 Db 460 gcagcagcagaagtccttttaaagggttgtaccnccagggaagaagatgcatcctccctt 519  
 |||||

Cp 1314 TCAGCAGCAGAA-GTCTTTAAAGGCTGTAC-ACCAGG-AAGAAAG-ATGCATCCTC-TT 1260  
 Db 520 ggccttggtggcaatcattttcc 542  
 |||||  
 Cp 1259 G-CCTTG-TGGCAATCATTTTCC 1239  
 |||||  
 RESULT 11  
 LOCUS AA152779 411 bp mRNA EST 11-FEB-1997  
 DEFINITION mr78c03.r1 Stratagene mouse heart (#937316) Mus musculus CDNA clone 603556 5' similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN ;  
 ACCESSION AA152779  
 NID 91724432  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 411)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouse@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:368988  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 207.  
 FEATURES  
 source  
 1..411  
 /organism="Mus musculus"  
 /strain="NIH/Swiss"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts. Average insert size: 1.0 kb; Uni-ZAP Xr Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"  
 /clone="603556"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /sex="pooled"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 <1..>411  
 BASE COUNT 75 a 122 c 131 g 83 t  
 ORIGIN  
 mRNA  
 Query Match 7.9%; Score 123; DB 50; Length 411;  
 Best Local Similarity 91.8%; Pred. No. 5.45e-134;  
 Matches 135; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 Db 264 ccagccggtgaccatgttcgacagacgcgctccctacgtggccctcgatgtattg 323  
 |||||  
 Qy 329 CCGGGCAGAGACCAATGTTTGACAGACCGGCTGCCGTACGTGCTCGATGCTCTG 388  
 |||||  
 Db 324 cgtgtgtgctggatggattgccttttgggaattcttacttcaaggcataccccccttcagcg 383  
 |||||  
 Qy 389 CGTGTGCTGGCTGGATTGCCCTTTTGCAATTTCTTACTTCAAGGCATACCCCTTCCAACG 448  
 |||||



Db	182	agaattttccgggttgggaagcccaactgacctctcttggtgattcccttccatccctctggc	240
Cp	858	AAAGAGTGGCCCTGAATAGAAAGACACACCTCTCTCTTAAGCTCTTTCGTATTCCTCGAC	799
Db	241	ataagagagtctctcatattggcattaatggagtggttgaacagctctcgggttataga	300
Cp	798	ATATGTAGTATTCATATGTAACCATCGCTGCAGTTGATTTTGACCAATCTGGATCACAA	739
Db	301	gatgccagaaaatg	313
Cp	738	CATCCAAGAAGATG	726
RESULT	15		
LOCUS	AA237956	313 bp	mRNA
DEFINITION	mx78q01.r1 Soares mouse NML Mus musculus cDNA clone 692496.5, similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN	EST	03-MAR-1997
ACCESSION	AA237956		
MID	91862038		
KEYWORDS	EST.		
SOURCE	house mouse		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 313)		
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information. MG1:426056 Possible reversed clone: similarity on wrong strand Seq primer: -28mi3 rev2 ET from Amersham High quality sequence stop: 57.		
FEATURES	Location/Qualifiers		
SOURCE	1..313 /organism="Mus musculus" /note="Vector: pT7T30-Pac (Pharmacia) with a modified polylinker. Site.1: Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTATCAATCTGAAGTGGAGGCGGCGGAATCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldi."		
BASE COUNT	83 a 58 c 85 g 87 t		
ORIGIN	/clone="692496" /clone_lib="Soares mouse NML" /tissue_type="Liver" /lab_host="DH10B" <1..>313		
Query Match	5.6%;	Score 88;	DB 66; Length 313;
Best Local Similarity	66.5%;	Pred. No. 6.59e-83;	
Matches	208; Conservative	0; Mismatches 102;	Indels 3; Gaps 3;

```
Db 3 gsgtccaactgtgacgtccactccagtggtgtttgtagtcagacactcgagaaagcc-ac 61
Cp 1037 GAGTCCAGTCAACACATCGCTCCAGTGGTGTATATATCAGAAATCGAGAAAGCCCCAC 978
Db 62 atatatggaanaagcaatgagggccaaactggagcaatggtcgtgaagaagcctggccaatt 121
Cp 977 ATAAATGGATACGGCAACAGACCAAAATTCAGTGTGGGGCGTAAGA-GTCTTGCCAGT 919
Db 122 tccctttaacttgagcttgaggttaagttgcgacaacagattgagttacttgggattg 181
Cp 918 CTCCTTTCATCTGCTGGCTTGAGATTAAGTCCACAAACAGCATGCAGTACATGGAAACG 859
Db 182 agaattttccyggttggaagccaaactggccttcttggatttccctt-caatcccttggc 240
Cp 858 AAGAGTGGCCTGAATAGAGGACAAACCTGCCTTCCCTAACTCTTTCGATTCCCTCGAC 799
Db 241 ataaggagttcttcattattggcatttaataatgagttgattttgaacagttcggttataga 300
Cp 798 ATATGTAGTATTCAATGTATACCATCGCTGCAGTTGATTTTGGACCAATCTGGATCACAA 739
Db 301 gatgccagaaatg 313
Cp 738 CATCCAGAGAGTG 726
```

Search completed: Thu Nov 6 10:41:38 1997  
Job time : 577 secs.







Matches 20; Conservative 27; Mismatches 32; Indels 4; Gaps 3;

Db 37 dhalkfifellyltrydlisrkipisalsvfv-ealevgysk--hknpynlmhaadvttq 93  
 QY 42 DESIKYPYKEDTIPVALLGGIIIPFSIIIVILGETLSV-YCNLLHSNFIIRNYIATIK 100  
 Db 94 tvhylllyktgvanwlteleifai 116  
 QY 101 AIGTEFLFGAAASQSLTDIAKYSI 123

RESULT 5  
 ID W11256 standard; Protein; 634 AA.  
 AC W11256;  
 DT 17-MAR-1997 (first entry)  
 DE Hippocampus calcium/calmodulin dependent phosphodiesterase #2.  
 KW Cyclic GMP stimulated phosphodiesterase; CGS-PDE; bovine; brain; cAMP;  
 KW adrenal gland; transmembrane signal; extracellular hormone; antibody;  
 KW neurotransmitter; cGMP; calcium/calmodulin dependent phosphodiesterase;  
 KW Cam-PDE; human.  
 OS Homo sapiens.  
 PN 03-DEC-1996.  
 PD 03-DEC-1996.  
 PF 19-APR-1991; 688356.  
 PR 19-APR-1991; US-688356.  
 PR 20-APR-1992; US-872644.  
 PR 29-AUG-1994; US-297494.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Beavo JA, Charbonneau H, Sonnenburg WK;  
 DR WPI; 97-033573/03.  
 DR N-PSDB; T51116.  
 PT DNA encoding cGMP-stimulated phosphodiesterase - for prodn. of  
 PT recombinant enzyme.  
 PS Example 7; Column 109-114; 68pp; English.  
 CC This sequence represents the calcium/calmodulin dependent  
 CC phosphodiesterase (Cam-PDE) encoded by the cDNA insert of pcamH3EF.  
 CC This cDNA insert was isolated from human hippocampus mRNA. The cyclic  
 CC nucleotide phosphodiesterases (PDEs) catalyze the hydrolysis of 3', 5',  
 CC cyclic nucleotides, such as cAMP and cGMP, to their corresponding  
 CC 5'-nucleotide monophosphates. The PDEs are therefore important in the  
 CC control of the cellular concentration of cyclic nucleotides. The PDEs  
 CC are, in turn, regulated by transmembrane signals or second messenger  
 CC ligands such as calcium ion or cGMP. The PDEs therefore have a central  
 CC role in regulating the flow of information from extracellular hormones,  
 CC neurotransmitters, or other signals that use the cyclic nucleotides as  
 CC messengers. PDEs are present in most of the cells and tissues of  
 CC eukaryotic organisms, but only in trace amounts. Cam-PDEs are responsive  
 CC to intracellular calcium, which leads to a decreased intracellular  
 CC concentration of cAMP, and/or cGMP. The cDNA encoding this sequence can  
 CC be used for the production of recombinant Cam-PDE, which may have  
 CC therapeutic and diagnostic uses. This protein may also be useful for  
 CC diagnostic antibody production.  
 SQ Sequence 634 AA.

Query Match 4.8%; Score 102; DB 20; Length 634;  
 Best Local Similarity 24.1%; Pred. No. 6.07e+00;  
 Matches 20; Conservative 27; Mismatches 32; Indels 4; Gaps 3;

Db 182 dhalkfifellyltrydlisrkipisalsvfv-ealevgysk--hknpynlmhaadvttq 238  
 QY 42 DESIKYPYKEDTIPVALLGGIIIPFSIIIVILGETLSV-YCNLLHSNFIIRNYIATIK 100  
 Db 239 tvhylllyktgvanwlteleifai 261  
 QY 101 AIGTEFLFGAAASQSLTDIAKYSI 123

RESULT 6  
 ID R69732 standard; Protein; 634 AA.  
 AC R69732;  
 DT 11-OCT-1995 (first entry)  
 DE Cyclic-GMP stimulated nucleotide PDE clone pcamH3EF.  
 KW Cyclic-GMP stimulated nucleotide phosphodiesterase; antibodies;

KW human; hormones; neurotransmitters; transmission regulation;  
 KW enzyme purification; clone pcamH3EF.  
 OS Homo sapiens.  
 PN 14-FEB-1995.  
 PD 14-FEB-1995.  
 PF 19-APR-1991; 688356.  
 PR 19-APR-1991; US-688356.  
 PR 20-APR-1992; US-872644.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Beavo JA, Charbonneau H, Sonnenburg WK;  
 DR WPI; 95-090205/12.  
 DR N-PSDB; Q83981.  
 PT New nucleic acid encoding cyclic-GMP stimulated nucleotide  
 PT phosphodiesterase - and related vectors and transformed cells,  
 PT useful for screening cpds. for phosphodiesterase modulating  
 PT activity.  
 PS Example 7; Columns 109-114; 69pp; English.  
 CC Q83981 encodes R69732 the human cyclic-GMP stimulated nucleotide  
 CC phosphodiesterase (Cam PDE) clone pcamH3EF. Eukaryotic cells  
 CC that express Cam PDE can be used to screen cpds. for the ability to  
 CC modulate Cam PDE activity. Cam PDEs are involved in regulating  
 CC the transmission of information from hormones, neurotransmitters  
 CC or other systems that use cyclic nucleotides as messengers.  
 CC Antibodies raised against Cam PDE can be used for enzyme purificn.,  
 CC or determination.  
 SQ Sequence 634 AA.

Query Match 4.8%; Score 102; DB 13; Length 634;  
 Best Local Similarity 24.1%; Pred. No. 6.07e+00;  
 Matches 20; Conservative 27; Mismatches 32; Indels 4; Gaps 3;

Db 182 dhalkfifellyltrydlisrkipisalsvfv-ealevgysk--hknpynlmhaadvttq 238  
 QY 42 DESIKYPYKEDTIPVALLGGIIIPFSIIIVILGETLSV-YCNLLHSNFIIRNYIATIK 100  
 Db 239 tvhylllyktgvanwlteleifai 261  
 QY 101 AIGTEFLFGAAASQSLTDIAKYSI 123

RESULT 7  
 ID R73055 standard; Protein; 1873 AA.  
 AC R73055;  
 DT 02-NOV-1995 (first entry)  
 DE Rabbit skeletal calcium channel (alpha)1-subunit.  
 KW Calcium channel; (alpha)1 subunit.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT Region 950...1100  
 FT /label= I187 monoclonal antibody epitope  
 FT Misc.difference 1808  
 FT /label= Thr  
 FT /note= "In Tanabe et al"  
 FT Misc.difference 1815  
 FT /label= Ala  
 FT /note= "In Tanabe et al"  
 FT Misc.difference 1835  
 FT /label= Ala  
 FT /note= "In Tanabe et al"  
 FT Modified\_site 79  
 FT /label= potential N-glycosylation site  
 FT Modified\_site 257  
 FT /label= see above  
 FT Modified\_site 797  
 FT /label= see above  
 FT Modified\_site 1464  
 FT /label= see above  
 FT Modified\_site 1674  
 FT /label= see above  
 FT Modified\_site 687  
 FT /label= potential cAMP-dependent phosphorylation  
 FT Modified\_site 1502  
 FT /label= see above







KW extracellular matrix; cytoskeleton; heterodimer; laminin receptor;  
KW immunoprecipitation; JAR; choriocarcinoma.  
OS Homo sapiens.

PH Key Location/Qualifiers  
FT Domain 1012..1037  
FT /label- Transmembrane\_Domain  
FT Peptide 1..23  
FT /label- Signal\_peptide  
FT Protein 24..1073  
FT /label- Mature\_protein  
FT Modified\_site\_223  
FT /label- Glycosylation\_site  
FT Modified\_site\_284  
FT /label- Glycosylation\_site  
FT Modified\_site\_370  
FT /label- Glycosylation\_site  
FT Modified\_site\_513  
FT /label- Glycosylation\_site  
FT Modified\_site\_731  
FT /label- Glycosylation\_site  
FT Modified\_site\_748  
FT /label- Glycosylation\_site  
FT Modified\_site\_891  
FT /label- Glycosylation\_site  
FT Modified\_site\_927  
FT /label- Glycosylation\_site  
FT Modified\_site\_958  
FT /label- Glycosylation\_site  
FT Binding\_site\_230..238  
FT /note= "Putative cation binding domain"  
FT Binding\_site\_324..332  
FT /note= "Putative cation binding domain"  
FT Binding\_site\_386..394  
FT /note= "Putative cation binding domain"  
FT Binding\_site\_441..449  
FT /note= "Putative cation binding domain"  
FT Domain 1040..1044  
FT /label- Cytoplasmic\_domain  
FT /note= "Conserved in virtually all integrins"  
PN W09219647-A.  
PD 12-NOV-1992.  
PF 27-APR-1992; U03527.  
PR 03-MAY-1991; US-695564.  
PA (SRI ) SCRIPPS RES INST.  
PI Quaranta V, Tamura RN;  
DR MPI; 92-398799/48.  
DR N-PSDB; Q31188.  
PT Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used  
PT for prodn. of antibodies and in detection of integrin sub-units  
PT in body samples  
PS Disclosure: Page 69-73: 115pp; English.  
CC The sequences given in R2821-22 are the human alpha 6A and 6B  
CC integrin subunits. Integrins are a family of cell surface receptors  
CC which serve cellular adhesion functions. These receptors form a link  
CC between the extracellular matrix and the cytoskeleton through their  
CC binding to various extracellular components. Each integrin receptor  
CC is a heterodimer comprised of an alpha and a beta subunit. Each alpha  
CC subunit tends to associate with only one type of beta subunit but  
CC there are several exceptions to this rule. These integrins correspond  
CC to the laminin receptor. The cytoplasmic domain of the 6A and 6B  
CC integrins differs from previously isolated alpha 6 integrins. The  
CC human alpha 6B was isolated from human choriocarcinoma cell line JAR  
CC by immunoprecipitation studies.  
SQ Sequence 1073 AA;

Query Match 4.4%; Score 94; DB 5; Length 1073;  
Best Local Similarity 25.3%; Pred No. 2, 21e+01;  
Matches 20; Conservative 23; Mismatches 29; Indels 7; Gaps 7;

Db 997 vrvtfpskt-vaqysgvpwailviallallvllwkcgffk-rnk-khdyat 1053  
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :  
Qy 207 VAVSIYVGLSRVSDYK-HHWSVDLTGLIQGALV-AILLVAV-YVSDFKERTSKERKEED 263

Db 1054 ybkacihagqpsdkerltsd 1072  
Qy 264 SHTT-LHETPTTGNHPSN 281  
Search completed: Tue Nov 4 10:20:24 1997  
Job time : 94 secs.

**This Page Blank (uspto)**

(TM)

Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

Tabular output not generated.

Sequence: 1 MFDKTRLPYVALDVLCVLLA.....HTTLHETPTTGNHYPSNHOP 284

Searched: 91006 seqs, 28888923 residues

Listing first 45 summaries

```
9:unann4 3:unann5 10:unann6 11:unann7 12:unann8
13:unann9 14:unann10 15:unenc 16:unrev
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	188	8.8	274	11	S69561	hypothetical protein	1.94e+11
2	181	8.5	289	11	S70114	hypothetical protein	1.86e+10
3	121	5.7	345	1	QXZL2M	NADH dehydrogenase (	1.26e-02
4	117	5.5	216	10	S32217	hypothetical protein	2.73e-02
5	110	5.2	437	9	H47070	hypothetical transmembr	2.37e-01
6	111	5.2	2220	13	A45290	calcium channel prot	1.83e-01
7	103	4.8	370	9	A45454	amine dehydrogenase	1.41e+00
8	102	4.8	1873	14	A30063	dihydropyridine rece	1.80e+00
9	102	4.8	2208	12	S37860	calcium channel prot	1.80e+00
10	99	4.7	157	13	S58012	probable olfactory r	3.76e+00
11	100	4.7	261	8	I38164	hypothetical protein	2.95e+00
12	99	4.7	342	16	S68129	NADH dehydrogenase (	3.76e+00
13	99	4.7	352	5	B34284	NADH dehydrogenase (	3.76e+00
14	99	4.7	368	9	S54809	1,6-nucleoyl-CoA desatu	3.76e+00
15	99	4.7	382	10	E64209	plipoprotein diacy	3.76e+00
16	100	4.7	388	4	GNWVDP	genome polyprotein -	2.95e+00
17	97	4.6	346	5	S10188	NADH dehydrogenase (	6.09e+00
18	97	4.6	366	10	B64413	hypothetical protein	6.09e+00
19	98	4.6	576	5	S54750	cycK protein - Rhizo	4.79e+00
20	97	4.6	723	10	D42707	H+/K+-transporting A	6.09e+00
21	97	4.6	885	8	A25817	nonstructural protei	6.09e+00

06-Sep-1996

06-Sep-1996

ACCESSIONS	S70114
REFERENCE	S70114
#authors	Fulton, L.
#submission	submitted to the EMBL Data Library, May 1996
#description	The sequence of S. cerevisiae cosmid 9819.
#accession	S70114
#molecule_type	DNA
#residues	1-289 #label FULL
#cross-references	EMBL:U51031
GENETICS	
#map_position	4R
SUMMARY	#length 289 #molecular-weight 33514 #checksum 4703
Query Match	8.5%; Score 181; DB 11; Length 289;
Best Local Similarity	27.0%; Pred. No. 1.86e-10;
Matches	67; Conservative 65; Mismatches 98; Indels 18; Gaps 16;
Db	38 qqperqfeyndltishyy--atte-rvnnmmifvfsvfpst-ill-i-qsil-adr-- 89
Qy	31 HTPFQGVCFNDESIKYPKEDTYPYALLGGIIPSIIVILGETLSVYNLLHSNFI 90
Db	90 rh-lfillytllslglawfstfftklfnwgrlpofldrcgvp-eglpdltlftak 147
Qy	91 RNNYIAIYKAIGTFLFGAASQSQTIAKYSIGRLRPHELDVCPDFDKSCINDGYIEY 150
Db	148 dvcttknherlidgfritcp-sghsesfagilyfwlcgqiltspimprkwmvaplp 206
Qy	151 YICRG-NAEVRKEG-RLSFYSGHSSFSMCMFLVALYLQARMKGDWARL-L-RTILOF-G 205
Db	207 llgaal-iastqtqdrhhfvdvilgmilgyimahahffyrifppiddplpfkplmd-dsd 264
Qy	206 LVAVSYIVGLSRVDYKHWSDVITGLIQALVAIIYAVVSDFFKERTSPKERKEEDSH 265
Db	265 vtleeavt 272
Qy	266 TTLHREPT 273
RESULT	3
ENTRY	OXXL2W #type complete
TITLE	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 -
ALTERNATE_NAMES	African clawed frog mitochondrion (SCGI)
ORGANISM	NADH-ubiquinone oxidoreductase chain 2 #formal_name mitochondrion Xenopus laevis #common_name African clawed frog
DATE	28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change 09-Sep-1994
ACCESSIONS	A00417
REFERENCE	A00155
#authors	Roer, B.A.; Ma, D.P.; Wilson, R.K.; Wong, J.F.H.
#journal	J. Biol. Chem. (1985) 260:9759-9774
#title	The complete nucleotide sequence of the Xenopus laevis mitochondrial genome.
#cross-references	MUID:85261388
#accession	A00417
#molecule_type	DNA
#residues	1-345 #label ROE
GENETICS	
#genome	mitochondrion
#genetic_code	SCGI
CLASSIFICATION	#superfamily NADH dehydrogenase (ubiquinone) chain 2
KEYWORDS	membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; oxidoreductase; respiratory chain
SUMMARY	#length 345 #molecular-weight 37671 #checksum 3982
Query Match	5.7%; Score 121; DB 1; Length 345;
Best Local Similarity	25.3%; Pred. No. 1.26e-02;
Matches	21; Conservative 27; Mismatches 28; Indels 7; Gaps 7;
Db	51 rateastkyfltaqaasal-llfsln-nawlgtewsiildtnplscatmtiaicmklg- 107
Qy	159 RVKEGRSLSYSGHSFSMYCMLFVALYLQARMKGDWARL-LRPTLOFGLVASYIVGLSR 217



```

TITLE      calcium channel protein type L - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
28-Apr-1995

ACCESSIONS A45290
REFERENCE   Soldatov, N.M.
#authors   Proc. Natl. Acad. Sci. U.S.A. (1992) 89:4628-4632
#journal   Molecular diversity of L-type Ca2+ channel transcripts in
#title     human fibroblasts.
#cross-references MUID:92262493
#accession A45290
#molecule_type mRNA
#residues  1-2220 #label SOL
#cross-references NCBI:105570; NCBI:105571
#experimental_source fibroblast
#note      this sequence is inconsistent with the nucleotide
          translation
#note      sequence extracted from NCBI backbone
COMMENT    This sequence was derived from multiple, partial, alternatively
          spliced mRNAs. The sequence shown is a composite representing the
          longest possible splice form. Two splice forms lead to premature
          stop codons.

FEATURE    #region alternatively spliced segment 1\
952-971    #region alternatively spliced segment 2\
1297-1324  #region alternatively spliced segment 3\
1325-1352  #region alternatively spliced segment 4\
1353-1363  #region alternatively spliced segment 5\
1864-1868  #region alternatively spliced segment 5\
SUMMARY    #length 2220 #molecular-weight 249069 #checksum 8928

Query Match      5.2%; Score 111; DB 13; Length 2220;
Best Local Similarity 23.2%; Pred. No. 1.83e-01;
Matches 29; Conservative 34; Mismatches 56; Indels 6; Gaps 6;

Db 902 ftnllfllllsllaedpvghtsfnrhlfydfvftftf-ialkllgnadyvf 960
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 9 YVALDVLCLLAGLFAILTS-RHFFPGVGY-FCNDESIRYKPTIPYALGGIIFP 66
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 961 tsiftleilkmtaygafihkgsfcrnyfnildllvsvslsfqgsaainvkvilrvl 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 -SIIVILGETLSVYCNLLHSNFIRNNYIATIVKAIGTFLFGAAASQSLTDIAKY-SIG 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1021 rvlrp 1025
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 R-LRP 128

RESULT 7
ENTRY   A44545 #type complete
TITLE   amine dehydrogenase (EC 1.4.99.3) heavy chain - Thiobacillus
          versutus
ORGANISM #formal_name Thiobacillus versutus
DATE     31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
31-Dec-1993
ACCESSIONS A44545
REFERENCE   Vellieux, F.M.D.; Kalk, K.H.; Drenth, J.; Hol, W.G.J.
#authors   Acta Crystallogr. (1990) B46:806-823.
#journal   Structure determination of quinoprotein methylamine
#title     dehydrogenase from Thiobacillus versutus.
#accession A44545
#status    preliminary
#molecule_type protein
#residues  1-370 #label VEL
KEYWORDS  oxidoreductase
SUMMARY   #length 370 #molecular-weight 35778 #checksum 3813

Query Match      4.8%; Score 103; DB 9; Length 370;
Best Local Similarity 31.3%; Pred. No. 1.41e+00;
Matches 25; Conservative 18; Mismatches 31; Indels 6; Gaps 6;

Db 59 lghslgaf-vilavaghsqdf-ala-stsfarsamgkrtqysqaieastfgasadsaa 115

```

```

QY 59 LGGIIPSIIVIIIGETLSVYCNLLHSNFIRNNY-IATIV-KAIGTFLFGAAASQSLT 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 avgqfsvgp-rdhlgncas 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 DIAKYSIGRLRPHFLDVCDP 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ENTRY   A30063 #type complete
TITLE   dihydropyridine receptor - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
          rabbit
DATE     30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
18-Jun-1993
ACCESSIONS A30063
REFERENCE   Tanabe, T.; Takeshima, H.; Mikami, A.; Flockerzi, V.;
#authors   Takahashi, H.; Kangawa, K.; Kojima, M.; Matsuo, H.; Hirose,
          T.; Numa, S.
#journal   Nature (1987) 328:313-318
#title     Primary structure of the receptor for calcium channel
          blockers from skeletal muscle.
#cross-references MUID:87258269
#accession A30063
#molecule_type mRNA
#residues  #length 1873 #label TAN
SUMMARY   #length 1873 #molecular-weight 212027 #checksum 5776

Query Match      4.8%; Score 102; DB 14; Length 1873;
Best Local Similarity 23.8%; Pred. No. 1.80e+00;
Matches 20; Conservative 22; Mismatches 39; Indels 3; Gaps 3;

Db 821 piraesvnrqilgydiaftsvftveivlmttygafihkgsfcrnyfnildllvavsl 880
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 48 PKREDTIPYALGGIIFP-SIIVILGETLSVYCNLLHSNFIRNNYIATIVKAIGTFL 106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 881 ismglestisvkvilrvlrp 904
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 FGAAASQSLTDIAKY-SIGR-LRP 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
ENTRY   A37860 #type complete
TITLE   calcium channel protein alpha-1 chain, skeletal muscle -
          common carp
ORGANISM #formal_name Cyprinus carpio #common_name common carp
DATE     31-May-1991 #sequence_revision 31-May-1991 #text_change
06-Dec-1996
ACCESSIONS A37860
REFERENCE   Grabner, M.; Friedrich, K.; Knaus, H.G.; Striessnig, J.;
#authors   Scheffauer, F.; Staudinger, R.; Koch, W.J.; Schwartz, A.;
          Glossmann, H.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1991) 88:727-731
#title     Calcium channels from Cyprinus carpio skeletal muscle.
#cross-references MUID:91126068
#accession A37860
#status    preliminary; nucleic acid sequence not shown; not
          compared with conceptual translation
#molecule_type mRNA
#residues  1-2208 #label GRA
#cross-references GB:M37203
SUMMARY   #length 2208 #molecular-weight 247817 #checksum 1915

Query Match      4.8%; Score 102; DB 12; Length 2208;
Best Local Similarity 26.5%; Pred. No. 1.80e+00;
Matches 22; Conservative 22; Mismatches 33; Indels 6; Gaps 5;

Db 968 sfrnkvlavadvftv-ftie-ivl--kmtvygafihkgsfcrnyfnildllvavsl 1023
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 48 PKREDTIPYALGGIIFP-SIIVILGETLSVYCNLLHSNFIRNNYIATIVKAIGTFL 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db   1024 smgnesstiasvklrvlvrp 1046
      : : | : | : | |||
Qy   108 GAAASOSLTIARIY-SIGR-LRP 128

RESULT    10
ENTRY     S58012         type fragment
TITLE     probable olfactory receptor tpor106 - human (fragment)
ORGANISM  #formal_name Homo sapiens #common_name man
DATE      13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change
          19-Apr-1996
ACCESSIONS S58012
REFERENCE   S57995
AUTHORS    Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.
SUBMISSION submitted to the EMBL Data Library, July 1995
DESCRIPTION Male germ cells from several mammalian species express a
              specific repertoire of olfactory receptor genes.
ACCESSION  S58012
STATUS     preliminary
MOLECULE_TYPE mRNA
RESIDUES   1-157 ##label VAN
CROSS_REFERENCES EMBL:X89667
SUMMARY    #length 157 #checksum 7167

Query Match           4.7%; Score 99; DB 13; Length 157;
Best Local Similarity 28.3%; Pred. No. 3.76e+00;
Matches 26; Conservative 26; Mismatches 31; Indels 9; Gaps 6;

Db   50 phff--cdlyalmkicstyintimhtegavvisgalafit--asya--ciliwvlr 103
      |||| | | ||| : : : : | : | : : | : | : | : | : | : | : | : |
Qy   128 PHFLVDCDP-DWSKINCSDGYIEYYICRGNAERKEGRLSFVGSHSSFSMYCMLEFVALYL 186

Db   104 ps-agkrwtfcsgshltvvai-fygtlsrw 133
      : ||| : : : : : : ||| : | : | : | : | : | : |
Qy   187 QARKKGDAWRLRPTQLQGLVAVSIVYGLSRV 218

RESULT    11
ENTRY     I38164         type complete
TITLE     hypothetical protein g - Pseudomonas sp.
ORGANISM  #formal_name Pseudomonas sp.
DATE      31-Jul-1992 #sequence_revision 31-Jul-1992 #text_change
          18-Jun-1993
ACCESSIONS I38164
REFERENCE   A38164
AUTHORS     Crouzet, J.; Levy-Schill, S.; Cameron, B.; Cauchois, L.;
            Rigault, S.; Rouyez, M.C.; Blanche, F.; Debussche, L.;
            Thibaut, D.
JOURNAL    J. Bacteriol. (1991) 173:6074-6087
TITLE      Nucleotide sequence and genetic analysis of a
            13.1-kilobase-pair Pseudomonas denitrificans DNA fragment
            containing five cob genes and identification of structural
            genes encoding cob(I)alamin adenosyltransferase, cobyric
            acid synthase, and bifunctional cobinamide
            kinase-cobinamide phosphate guanylyltransferase.
CROSS_REFERENCES MIMD:92011366
ACCESSION  I38164
STATUS     preliminary
MOLECULE_TYPE DNA
RESIDUES   1-261 ##label CRO
CROSS_REFERENCES GB:M62866
SUMMARY    #length 261 #molecular_weight 27374 #checksum 8744

Query Match           4.7%; Score 100; DB 8; Length 261;
Best Local Similarity 31.1%; Pred. No. 2.95e+00;
Matches 19; Conservative 18; Mismatches 21; Indels 3; Gaps 3;

Db   102 vlkalipflialtalalgakpmgdvdq-srvtfvtlltvp-ligfydgvgspgts 159
      : : : | | : | : | : | : | : | : | : | : | : | : | : | : |
Qy   197 LLRPTLQGLVAVSIVUGL-SRVSDDKKHWSVDLTLOALVAILAVAYVSDFDKERTS 255

Db   160 f 160

```

```

GENETICS          shown
#gene             ND2
#genome           mitochondrial
#genetic_code     SGC8
#superfamily      NADH dehydrogenase (ubiquinone) chain 2
#CLASSIFICATION   membrane-associated complex; mitochondrion; NAD;
#KEYWORDS          oxidoreductase
SUMMARY           #length 352 #molecular-weight 39009 #checksum 8650

Query Match      4.7%; Score 99; DB 5; Length 352;
Best Local Similarity 32.6%; Pred. No. 3.76e+00;
Matches 28; Conservative 22; Mismatches 26; Indels 10; Gaps 9;

Db 51 rnieatkyflvq-fsaallnsal-igawfsgwsali-pmesfscisvalafnlg 107
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
QY 159 RVKEGLSFYSGHSFSMYCMFLFVALYQARMKGDNARLLRPTLQF-GL-VAVSIIYVGLS 216
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
Db 108 -laac-hfwlpdvlgqpfqgliia 131
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
QY 217 RVSDYKHHW-SDVETGL--IQGALVA 239

RESULT 14
ENTRY      S54809 #type complete
TITLE      linoleoyl-CoA desaturase (EC 1.14.99.25) - Spirulina
            platensis
ALTERNATE_NAMES Deltac6-desaturase
ORGANISM      #formal_name Spirulina platensis
DATE          08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
            23-Feb-1996
ACCESSIONS   S54809
REFERENCE     Tasaka, Y.
#submission  submitted to the EMBL Data Library, May 1995
#accession   S54809
#status      preliminary
#molecule_type DNA
#residues    1-368 #label TAS
#cross-references EMBL:X87094
KEYWORDS     oxidoreductase
SUMMARY      #length 368 #molecular-weight 42083 #checksum 6298

Query Match      4.7%; Score 99; DB 9; Length 368;
Best Local Similarity 31.9%; Pred. No. 3.76e+00;
Matches 15; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 67 wmkllqciqlvgfsavafnshdngbgyskygwvnylsglthdai 113
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
QY 194 WARLLRPT-LQFGLVAVSIYVGL-SRVSDY-KHWSVDLTGLIQGAL 237

RESULT 15
ENTRY      E64209 #type complete
TITLE      prolipoprotein diacylglycerol transferase homolog -
            Mycoplasma genitalium (SGC3)
ORGANISM    #formal_name Mycoplasma genitalium
DATE        17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
            17-Nov-1995
ACCESSIONS   E64209
REFERENCE     Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
            R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
            Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
            Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
            Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
            J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
            Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
            C.A.; Venter, J.C.
            Science (1995) 270:397-403
            The minimal gene complement of Mycoplasma genitalium.
ACCESSION    E64209
#status      preliminary

```

```

#molecule_type DNA
#residues        1-382 #label TIGR
#cross-references GB:L43967; TIGR:MG086
#experimental_source strain G-37
#note            neither nucleotide sequence nor conceptual translation
            is shown
GENETICS
#genetic_code     SGC3
SUMMARY           #length 382 #molecular-weight 44431 #checksum 8286

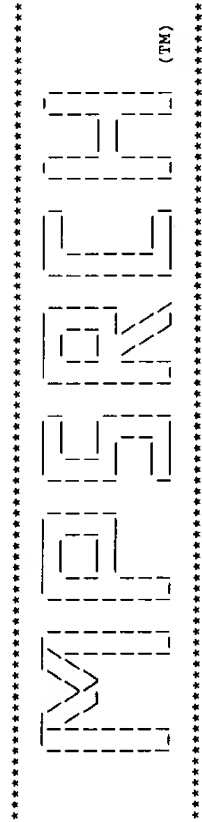
Query Match      4.7%; Score 99; DB 10; Length 382;
Best Local Similarity 34.7%; Pred. No. 3.76e+00;
Matches 17; Conservative 13; Mismatches 18; Indels 1; Gaps 1;

Db 22 gliysigilfailmfvfkliycykldgnsfyffifiafvtmvlgarlws 70
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |
QY 61 GLIIPFSIIIVILGETLS-VYCNLLHSNSEFIRNNYIATYKAIGTFLEG 108
| | : : : : : | | : : : : : | | : : : : : | | : : : : : |

Search completed: Tue Nov 4 10:18:32 1997
Job time : 65 secs.

```

**This Page Blank (uspto)**



Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Nov 4 10:16:22 1997; Maspar time 7.79 Seconds  
773.169 Million cell updates/sec

Tabular output not generated.

Title: >US-08-842-827-2  
Description: (1-284) from US08842827.pep  
Perfect Score: 2125  
Sequence: 1 MFQKTRLPYVALDVLCVLLA.....HTTLHETPTTGNHYPNSHQ 284

Scoring table: PAM 150  
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 47.863; Variance 91.115; scale 0.525

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match Length	ID	Description	Pred. No.		
1	443	20.8	341 11	YSX3_CABEL	HYPOTHETICAL 39.0 KD	8.02e-64	
2	121	5.7	345 7	NU2M_XENLA	NADH-UBIQUINONE OXIDO	8.14e-04	
3	105	4.9	345 11	YDE9_SCHPO	PUTATIVE MITOCHONDRIA	1.30e-01	
4	103	4.8	370 3	DHXL_THIVE	METHYLAMINE DEHYDROGE	2.36e-01	
5	102	4.8	1852 2	CIC1_CYPCA	DIHYDROPYRIDINE-SENS	3.18e-01	
6	102	4.8	1873 2	CIC1_RABIT	DIHYDROPYRIDINE-SENS	3.18e-01	
7	100	4.7	261 11	YC89_PSEDE	HYPOTHETICAL 27.4 KD	5.72e-01	
8	99	4.7	352 7	NU2M_PARLI	NADH-UBIQUINONE OXIDO	7.65e-01	
9	99	4.7	382 6	LGT_MYCGE	PUTATIVE PROLIPOPROTE	7.65e-01	
10	100	4.7	3388 7	POLG_DEN2P	GENOME POLYPROTEIN (C	5.72e-01	
11	98	4.6	346 11	YQCL_BACSU	HYPOTHETICAL 38.3 KD	1.02e+00	
12	97	4.6	346 6	NU2M_CHICK	NADH-UBIQUINONE OXIDO	1.36e+00	
13	97	4.6	443 6	MDM1_YEAST	STRUCTURAL PROTEIN MD	1.36e+00	
14	98	4.6	676 2	CCME_RHIME	CYTOCHROME C-TYPE BIO	1.02e+00	
15	98	4.6	704 7	PM2A_LYCES	PLASMA MEMBRANE ATPAS	1.02e+00	
16	97	4.6	723 2	CADA_BACFI	PROBABLE CADMIUM-TRAN	1.36e+00	
17	98	4.6	3391 7	POLG_DEN2P	GENOME POLYPROTEIN (C	1.02e+00	
18	98	4.6	3391 7	POLG_DEN2N	GENOME POLYPROTEIN (C	1.02e+00	
19	97	4.6	3391 7	SYFA_ECOLI	PHENYLALANYL-TRNA SYN	2.39e+00	
20	95	4.5	327 9	NU2M_COTJA	NADH-UBIQUINONE OXIDO	2.39e+00	
21	95	4.5	346 6	NU2M_COTJA	NADH-UBIQUINONE OXIDO	2.39e+00	
22	96	4.5	347 7	NU2M_PROVI	NADH-UBIQUINONE OXIDO	1.80e+00	

ALIGNMENTS									
RESULT	1	2	3	4	5	6	7	8	9
ID	YSX3_CABEL	STANDARD	PRT	341	AA				
AC	Q10022								
DT	01-OCT-1996	(REL. 34, CREATED)							
DT	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)							
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)							
DE	HYPOTHETICAL 39.0 KD PROTEIN T28D9.3 IN CHROMOSOME II.								
GN	T28D9.3								
OS	CAENORHABDITIS ELEGANS.								
OC	EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN-BRISTOL N2;								
RA	WATERSTON R.								
RL	SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.								
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).								
CC	-1- SIMILARITY: WEAK, TO YEAST D9719.9.								
DR	EMBL; U28738; G861267; -								
DR	WORMPEP; T28D9.3; CE02068.								
KW	HYPOTHETICAL PROTEIN; TRANSMEMBRANE.								
FT	TRANSMEM 30 50	POTENTIAL.							
FT	TRANSMEM 71 91	POTENTIAL.							
FT	TRANSMEM 122 142	POTENTIAL.							
FT	TRANSMEM 223 243	POTENTIAL.							
FT	TRANSMEM 257 277	POTENTIAL.							
SQ	SEQUENCE 341 AA; 89AE6E81 CRC32;								
Query Match 20.8%; Score 443; DB 11; Length 341;									
Best Local Similarity 34.1%; Pred. No. 8.02e-64;									
Matches 79; Conservative 61; Mismatches 80; Indels 12; Gaps 12;									
Db	52	grffcdsdriryrkdtitavqlmlylnvlnaattvlfveyymqkvesni-nnprrw	110						
QY	35	QRGVFCNDSIRYPKEDTI-PYAL-LGGIIPFISIIILGETL-SVYCNLLHNSFI-	90						
Db	111	rnshllvfrlltlyfysqigfvmnlvltkvgrlphfldvcklandtcvtgds	170						
QY	91	RNNVIATV-KAIGFLFGAAA-SOSLI-DI-ANYSIGRLRPHFLDVCDDPWSK-INC-S	144						
Db	171	hrytdyctgppelvlearksfygsghsavlcatwsalyiqarlgpvlmnnrvpisp	230						
QY	145	DGVIETIYICRGAERKGLSFYSFGHSFMYCMFLFVLYLQARMKGDWA-KLLRPTLQ	203						
Db	231	tlmfmglgtsfricdnkhwsdvlfvfigiflavtyctfctdflfsnst	282						
QY	204	FLGVAVSIYVGLSRVSDYKHHSDVLTGLIOGALVALVAVVYVSDFFKRTS	255						





```
FT CONFLICT 1808 1808 T -> M (IN REF. 2).
FT CONFLICT 1815 1815 A -> V (IN REF. 2).
FT CONFLICT 1835 1835 A -> E (IN REF. 2).
SQ SEQUENCE 1873 AA; 212028 MW; BA27142C CRC32;

Query Match 4.8%; Score 102; DB 2; Length 1873;
Best Local Similarity 23.8%; Pred. No. 3.18e-01;
Matches 20; Conservative 22; Mismatches 39; Indels 3; Gaps 3;

Db 821 piraesvnrqilgfydiafsvtveivkmtygafllkgsfcrnyfnldllvavsl 880
QY 48 PKYEDTIPYALGGIIPF-SIIVILGETLSVYCNLLHSNFIRNVIATYKAIGTFL 106
Db 881 ismglesstisvkvilvrlvp 904
QY 107 FGAASQSLTDIAKY-SIGR-LRP 128

RESULT 7
ID YC89_PSEDE STANDARD; PRT; 261 AA.
AC P29942;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DE PUTATIVE PROLIPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-).
OS PSEUDOMONAS DENITRIFICANS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
CC PSEUDOMONADACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92011366.
RA CROUZET J., LEVY-SCHIL S., CAMERON B., CAUCHOIS L., RIGAUD S.,
RA ROUVEZ M.-C., BLANCHE F., DEBUSSCHE L., THIBAUT D.;
RL J. BACTERIOL. 173:6074-6087(1991).
CC -!- SUBCELLULAR LOCATION: MEMBRANE (PROBABLE).
DR EMBL; M62866; G551929; -.
DR PIR; I38164; I38164.
KW HYPOTHETICAL PROTEIN; MEMBRANE.
SQ SEQUENCE 261 AA; 27374 MW; 5D508436 CRC32;

Query Match 4.7%; Score 100; DB 11; Length 261;
Best Local Similarity 31.1%; Pred. No. 5.72e-01;
Matches 19; Conservative 18; Mismatches 21; Indels 3; Gaps 3;

Db 102 vkalpfillatalfgkpnmgvdqgh-srvtpfvflltvp-llgfydvgfpgtgts 159
QY 197 LLRPTLQFLGVAISIYVGL-SRVSDYKHHWSVLTGLIQGALVAILVAVIYVSDFFKRTS 255
Db 160 f 160
QY 256 F 256

RESULT 8
ID NU2M_PARLI STANDARD; PRT; 352 AA.
AC P12771;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 (EC 1.6.5.3).
GN ND2.
OS PARACENTROTUS LIVIDUS (COMMON SEA URCHIN).
OC MITOCHONDRION.
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
CC EUCHINOIDEA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 8929181.
RA CANTATORE P., ROBERTI M., RAINALDI G., GADALETA M.N., SACCONI C.;
RL J. BIOL. CHEM. 264:10965-10975(1989).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; J04815; G453499; -.
DR PIR; B34284; B34284.

KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 352 AA; 39009 MW; 39BB4327 CRC32;

Query Match 4.7%; Score 99; DB 7; Length 352;
Best Local Similarity 32.6%; Pred. No. 7.65e-01;
Matches 28; Conservative 22; Mismatches 26; Indels 10; Gaps 9;

Db 51 rniatkiyflvga-fsaallnsal-icawfsgwsali-pmesfslslvalafnlq 107
QY 159 RVKGRUSFSGHSSFSMYCMLFVALYIQARKMGDMARLPTLQF-GL-VAVSIYVGLS 216
Db 108 -laec-hfwipdvlgqlpfqglila 131
QY 217 RVSDYKHHW-SDVLTGL--IOGALVA 239

RESULT 9
ID LGT_MYCGE STANDARD; PRT; 382 AA.
AC P47332;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE PUTATIVE PROLIPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-).
GN LGT OR MG086.
OS MYOPLASMA GENITALIUM.
OC MYOPLASMA.
OC PROKARYOTA; TENERICUTES; MOLLICUTES; MYCOPLASMA; MYCOPLASMALES;
CC MYCOPLASMATACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITCHMAN J.D., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RL SCIENCE 270:397-403(1995).
CC -!- FUNCTION: TRANSFERS THE N-ACYL DIGLYCERIDE GROUP ON WHAT WILL
CC BECOME THE N-TERMINAL CYSTEINE OF MEMBRANE LIPOPROTEINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- PATHWAY: FIRST STEP IN LIPOPROTEIN BIOSYNTHESIS.
DR EMBL; U39688; GI045763; -.
FT TRANSFERRASE; TRANSMEMBRANE.
FT TRANSFERRASE 18 38 POTENTIAL.
FT TRANSFERRASE 53 73 POTENTIAL.
FT TRANSFERRASE 92 112 POTENTIAL.
FT TRANSFERRASE 141 161 POTENTIAL.
FT TRANSFERRASE 213 233 POTENTIAL.
FT TRANSFERRASE 243 263 POTENTIAL.
FT TRANSFERRASE 274 294 POTENTIAL.
FT TRANSFERRASE 302 322 POTENTIAL.
SQ SEQUENCE 382 AA; 44431 MW; 5D374E0F CRC32;

Query Match 4.7%; Score 99; DB 6; Length 382;
Best Local Similarity 34.7%; Pred. No. 7.65e-01;
Matches 17; Conservative 13; Mismatches 18; Indels 1; Gaps 1;

Db 22 glivsigilfailmfvfklycyklqdnfsyffiflavltmvlglarls 70
QY 61 GIIPFSIIVILGETLS-VYCNLLHSNFIRNVIATYKAIGTFLFG 108

RESULT 10
ID POLG_DEN2P STANDARD; PRT; 3388 AA.
AC P12823;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E;
DE NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3);
```



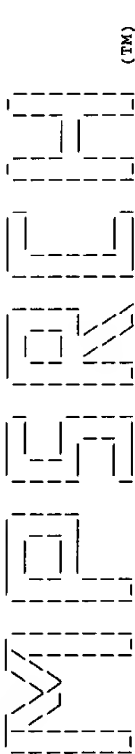




Search completed: Tue Nov 4 10:17:09 1997  
Job time : 47 secs.

**This Page Blank (uspto)**

\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 10:42:02 1997; Maspar time 1306.58 Seconds  
Tabular output not generated.  
1316.341 Million cell updates/sec

Title: >US-08-842-827-3  
Description: (1-1566) from US08842827.seq  
Perfect Score: 1566  
N.A. Sequence: 1 CCTGTGGGAGAGACGCCGCGG.....CGAAGAAAAA.....AAAAA 1566  
Comp: GGACACCTCTCTCGCGGCC.....GGTTTTTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-new3  
1: BCT2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database: genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
Database: genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL  
Database: u-emb150\_99  
122: part1

Statistics: Mean 11.944; Variance 6.199; scale 1.927

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.	
1	429	27.4	1212	85	D84376	House mouse; Musculus	0.00e+00	
2	265	16.9	1490	88	MUSHP1P	Mus musculus (clone H	4.61e-178	
C	3	130	3353	82	HUMORFA02	Human mRNA for KIAA00	7.06e-72	
4	111	7.1	2206	91	RNDRI42	R.norvegicus mRNA for	1.76e-57	
5	102	6.5	1444	78	HSU79294	Human clone 23748 mRN	9.39e-51	
C	6	41	215	57	I28278	Sequence 5 from paten	2.48e-08	
7	34	2.2	215	57	I28278	Sequence 5 from paten	3.13e-04	
8	32	2.0	1727	45	SSM27R	S.solidissima mRNA fo	4.01e-03	
C	9	31	2.0	4601	40	DMU11584	Drosophila melanogast	1.39e-02
C	10	31	2.0	14626	35	BMOFLCEG	Bombix mori fibroin 1	1.39e-02
C	11	31	2.0	19517	40	DMU37541	Drosophila melanogast	1.39e-02
12	30	1.9	1427	39	DBAAC3	Dictyostelium discoide	4.73e-02	
13	29	1.9	1512	46	BBCA	B.taurus mRNA for cyc	1.57e-01	
14	29	1.9	1837	91	RNF2XMR	R.norvegicus P2X mRNA	1.57e-01	
15	30	1.9	1862	64	PSMANC	Pyromyces sp. mRNA fo	4.73e-02	
16	29	1.9	8567	16	CAGNAT2	C.aureus GNAT2 gene	1.57e-01	
C	17	28	1.8	363	91	RNA7CINT	R.norvegicus mRNA for	5.09e-01
C	18	28	1.8	417	40	DOATRICH	D.obscura A-T-rich re	5.09e-01
19	28	1.8	793	122	NMD430	Mouse Murr1 mRNA, exo	5.09e-01	
20	28	1.8	793	118	D85430	Mouse Murr1 mRNA, exo	5.09e-01	
21	28	1.8	793	14	MMD430	Mouse Murr1 mRNA, exo	5.09e-01	
22	28	1.8	1524	43	LPACT3	L.polyphemus mRNA for	5.09e-01	
23	28	1.8	1737	39	DL19	Dictyostelium discoide	5.09e-01	
24	28	1.8	2049	44	S67872	Zw-glucose-6-phosphat	5.09e-01	
C	25	28	1.8	2630	43	PCU53325	Plasmodium chabaudi d	5.09e-01
26	28	1.8	3077	44	PFU07706	Plasmodium falciparum	5.09e-01	
27	28	1.8	3774	91	RNIAC	R.norvegicus mRNA for	5.09e-01	
28	28	1.8	6372	39	DIICNPA	Dictyoselium discoide	5.09e-01	
29	28	1.8	10359	44	PFARPI	Plasmodium falciparum	5.09e-01	
C	30	28	1.8	192274	110	HS435C23	Human DNA sequence **	5.09e-01
C	31	28	1.7	347	54	A23327	O.sativa mRNA for T23	1.61e+00
32	27	1.7	451	60	BNU59459	Brassica napus antifu	1.61e+00	
33	27	1.7	825	65	SCCOX1	Yeast cytochrome c ox	1.61e+00	
C	34	27	1.7	1095	61	DDICSA	Dictyostelium discoide	1.61e+00
35	27	1.7	1128	43	MPU15677	Myrmecia pilosula H18	1.61e+00	
36	27	1.7	1381	39	CTHRP45MR	C.tentans mRNA for hr	1.61e+00	
37	27	1.7	1702	43	PFAGPTA	Plasmodium falciparum	1.61e+00	
C	38	27	1.7	1803	73	HSARGBPIA	H.sapiens mRNA for Ar	1.61e+00
39	27	1.7	2035	39	DDU53884	Dictyostelium discoide	1.61e+00	
40	27	1.7	2045	42	PNLGSYN	Panulirus argus gluta	1.61e+00	
C	41	27	1.7	2798	42	HYDHP701B	Hydra magnipapillata	1.61e+00
42	27	1.7	3700	39	DGDP80G	D.discoideum gp80 gen	1.61e+00	
C	43	27	1.7	7989	44	PFU31083	Plasmodium falciparum	1.61e+00
44	27	1.7	7989	44	PFU31083	Plasmodium falciparum	1.61e+00	
45	27	1.7	9353	44	PFU67959	Plasmodium falciparum	1.61e+00	

ALIGNMENTS

RESULT	1	D84376	1212 bp	mRNA	ROD	03-OCT-1996
LOCUS		House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid phosphatase, complete cds.				
ACCESSION		D84376				
NID		91487872				
KEYWORDS		Phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.				
SOURCE		Mus musculus				
ORGANISM		Mus musculus				
		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
REFERENCE		1	(bases 1 to 1212)			
AUTHORS		Kanoh, H.				
TITLE		Direct Submission				
JOURNAL		Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Hideo Kanoh, Sapporo Medical University School of Medicine, Department of Biochemistry, West-17, South-1, Sapporo, Hokkaido 060, Japan (E-mail: kanoh@serpent.cc.sapmed.ac.jp, Tel: 011-611-2111(ex.2290), Fax: 011-612-5861)				

```

REFERENCE 2 (bases 1 to 1212)
AUTHORS Kai,M., Wada,I., Imai,S., Sakane,F. and Kanoh,H.
TITLE Identification and cDNA cloning of 35-kDa phosphatidic acid
phosphatase (type 2) bound to plasma membranes. Polymerase chain
reaction amplification of mouse H202-inducible hic53 clone yielded
the cDNA encoding phosphatidic acid phosphatase
J. Biol. Chem. 271 (31), 18931-18938 (1996)
JOURNAL 96324980
MEDLINE Location/Qualifiers
FEATURES 1..1212
Source /organism="Mus musculus"
287..1138 /tissue_type="kidney"
CDS /evidence=experimental
/codon_start=1
/product="Phosphatidic acid phosphatase"
/db_xref="PID:G148783"
/translation="MFKRLRFLVVALDVICVLGLFPAILTSRHPPFRGIFCNDDS
IKPIEDIPALGGIVPECIIVNSIGSLSYFNVLSHSFVGNPYATIKAV
GALFGVSQSUTDIATKISLPHFLATCNPDWKSINCSGDIYICOGNEEKV
KEGLFYSQHSFSNYCMFLVLIQARMGWRLRPLMQLFGLIAFYVGLSRV
SDYKHMSDVTGLIOGAAMAILVALYVSDFKDTSHYKERKEDPHPTLTHETASRN
YSTNHEP"
BASE COUNT 277 a 316 c 313 g 306 t
ORIGIN
Query Match 27.4%; Score 429; DB 85; Length 1212;
Best Local Similarity 76.8%; Pred. No. 0.00e+00;
Matches 636; Conservative 0; Mismatches 189; Indels 3; Gaps 2;
Db 274 ccagcgggtgaccatgttcacaaagacgqcgctgcgtacgtggccctcgatgtgattg 333
|||
Qy 329 CCGGCGAGACACCATGTTGACRAGACGGCGTCCGCTAGCTGGCCCTCGATGTCCTG 388
|||
Db 334 cgtgtgtcgtcgtgattgctcttttgaactcttacttcaagga--taccoc--ttcca 390
|||
Qy 389 CGTGTGTGCTGCTCCATGCTATGGCTGTGTTCTAAATTTGGGCCAAATATATCCATTCA 448
|||
Db 391 gcgaggaatatctgttaagtactcaccatcaagtacccttcaaggaagacacaccatcc 450
|||
Qy 449 GAGAGGCTTTTCTGTAAGACACACACATCACTACCTACCTACCTACCTACCTACCT 508
|||
Db 451 ttatgccttatagtggaatagtcattccattcctgtatctgttatgattgagaga 510
|||
Qy 509 ATCCACTGCTCTATCCATAGTGGGGTGGCTTGTGCCCGTTCTCTATATTTCTTGGAGA 568
|||
Db 511 atctctgtctgttactttaaagtcttgcattcgaaattccttctgcggaaatccctacat 570
|||
Qy 569 AACCTGTCTCTTACTGTAACTTTTGACCTCAATTCCTTATCAGTAATACTACAT 628
|||
Db 571 agccacatttacaagcgtcgagaccctttgttcgagtcctcagctagtcagtcctt 630
|||
Qy 629 AGCCACTATTTTACAAAGCCATTGGAACCTTTTATTTGTGTCAGCTGTACTAGTCCCT 688
|||
Db 631 gactgacatcgttaagtactactaaggcagtttgcggcgccactcttggcttatctgtaa 690
|||
Qy 689 GACTGACATTGCCAAGATTCAATAGGACAGCTGCGCGCTCACTTCTGGATGTTGGA 748
|||
Db 691 cccagactggtcacaataatcaactgcagtgatggcttatattgaggactacatatgtcaagg 750
|||
Qy 749 TCCAGATTGGTCAAAAATCAACTGTCAGCGATGTTTACATTGAATCACTACATATGTCGAGG 808
|||
Db 751 gaatgaagagaaagtcaagggagggcaggttcttcttactcgggacacactcttcattctc 810
|||
Qy 809 GAATGCAAAAGAGTTAAGGAAGGAGGTTGCTTCTATTACGCGCCACTCTCTGTTTC 868
|||
Db 811 tatgtactcgtatgcttctgtcgaacttatcttcaagccagatgaaggagagactgggc 870
|||
Qy 869 CATGTACTGTCATGCTGTTGTGGCAGCTTAATCTCAAGCCAGGATGAAGGGAGACTGGGC 928
|||
Db 871 aagactcttaagcccatgctccagttttgggctcattgtcttttccatatatgtgggct 930
|||
Qy 929 AAGACTCTTACGCCCCCACTGCAATTTGGTCTTGTGTGCGGTATCCATTATGTGGGCT 988
|||

```

---

```

Db 931 ttctcgagtgctgactacaaacacacactggagtgacgtcacagttggactcattcagg 990
|||
Qy 989 TTCTCGAGTTTCTGATTATAACACCACTGGAGCGATGTGTGACTGGACTCAITTCAGGG 1048
|||
Db 991 agctgctatgctactcgttctgttctgtatgtatccgattcttccaaggacacacattc 1050
|||
Qy 1049 AGCTCTGGTTCATATTAGTCTGTATGTATGTATGTCGGATTCTTCAAGAAGAATTC 1108
|||
Db 1051 ttacaagagagaaagagagaggtacccacacacgactctccatgaac 1098
|||
Qy 1109 TTTTAAAGAAAGAAAGAGGAGGACTCTCATACAACCTCTGCATGAAC 1156
|||
RESULT 2
LOCUS MUSHPIP 1490 bp mRNA ROD 11-JAN-1996
DEFINITION Mus musculus (clone HIC-53) hydrogen peroxide-inducible protein
mRNA, complete cds.
ACCESSION L43371
NID g1161099
KEYWORDS hydrogen peroxide; hydrogen peroxide-inducible protein.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 1490)
AUTHORS Egawa,K., Yoshiwara,M., Shibamura,M. and Nose,K.
TITLE Isolation of a novel ras-recision gene that is induced by hydrogen
peroxide from a mouse osteoblastic cell line, MC3T3-E1
FEBS Lett. 372 (1), 74-77 (1995)
JOURNAL 96032549
MEDLINE Location/Qualifiers
FEATURES 1..1490
Source /organism="Mus musculus"
/clone="HIC-53"
/cell_line="MC3T3-E1"
/cell_type="osteoblast"
/dev_stage="embryo"
/tissue_type="calvarilag"
<..978
/codon_start=1
/product="hydrogen peroxide-inducible protein"
/db_xref="PID:g1161100"
/translation="ARASLELARPVGEGTGORRPIGRDEPALREQAPEAQIAASTVVK
ESRKSQVRRPIOTGPGCSQSKRLAVAMGRRPBGAGLGYLDRLGLFLPPLAP
GGDTQPTWMDKTRLPYVALDVICVLGLPAILTSRHPTFORIFCNDSIKPY
KEDTIPYALGGIVTFPIIIVHSIGESLSYFNVLSHSFVGNPYATIKAVGAPLV
GVSQSUTDIATKISLPHFLATCNPDWKSINCSGDIYICOGNEEKVKEGR
SFSGTFLDYLVHAVCTLSRRLRLRPLMQLFGLIAFYVGLFSSV"
mRNA <..1490
3'UTR 979..1490
BASE COUNT 362 a 367 c 368 g 393 t
ORIGIN
Query Match 16.9%; Score 265; DB 88; Length 1490;
Best Local Similarity 75.1%; Pred. No. 4.61e-178;
Matches 434; Conservative 0; Mismatches 139; Indels 5; Gaps 3;
Db 321 ccagccggtgaccatgttcacaaagacgqcgctgcgtacgtggccctcgatgtgattg 380
|||
Qy 329 CCGGCGAGACACCATGTTTGACAAGACGGCGTCCGCTACGTGGCCCTCGATGTCCTG 388
|||
Db 381 cgtttactcgtcgtgattgctcttttgaactcttacttcaaggca--taccoc--ttcca 437
|||
Qy 389 CGTGTGCTGCTTCCATGCTTATGGCTGTCTTCTAAATTTGGCCAAATATATCCATTTC 448
|||
Db 438 gcgaggaatatcttctgaatgatgactccatcaagtacccttcaaggaagacacacatacc 497
|||
Qy 449 GAGAGGCTTTTCTGTAAGACAAACAGCATCAACTATCCGTACCATGACAGTACCGCCG 508
|||
Db 498 ttaagccttattagggtggaatagtcattccattcctgtattatcgttatgattgagaga 557
|||

```







AUTHORS Hunt,T. and Standart,N.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-1990) Hunt T., Standart N., Department of Biochemistry, Tennis Court Road, Cambridge, CB2 1QW, England  
REFERENCE 2 (bases 1 to 1727)  
AUTHORS Standart,N., Dale,M., Stewart,E. and Hunt,T.  
JOURNAL Genes Dev. 4 (1990). In press  
COMMENT Forms a complex with the product of the large M1 subunit of ribonucleotide reductase

FEATURES  
source location/Qualifiers  
1..1727  
/product="M2 (small), subunit of ribonucleotide reductase"  
/dev\_stage="oocyte"  
/cell\_type="oocyte"  
96..1250  
/EC\_number="1.17.4.1"  
/coo\_start=1  
/product="M2 (small), subunit of ribonucleotide reductase"  
/db\_xref="PID:g10342"  
/db\_translation="MUSITRKEELNSGNLKMITEENPKVKVLGEINFORSTOK  
TPKQEIKPVYKSQVEFLADNPRFVLPFIQYHDIMWKYKKAESAFTTAEVDLS  
TAHAWMSLKKEKHIEISHVLAFFAASDGIVENILVERFSKEOVTEARGCYGFOIAM  
ENTHSEMYSLIIDITFKDPQRDFLFNAIETMPCKVKADAMWINDDSSSYAVNRV  
AFAAVEIGSFSGIFPWKKRGIMPGLTFSNELISRDGLHCDFACLMFSLHVNRV  
SOERHIIDIAEVKIPOVELFALPCRLIGMNCIDLMRQYIEFVADRLLLELKCDKLYN  
KENPFPMEHISELGNTFEFRVGEYQKMGVMGGNGTGDSHAFTILDADF"

polyA\_signal  
BASE COUNT 586 a 260 c 324 g 557 t  
ORIGIN

Query Match 2.0%; Score 32; DB 45; Length 1727;  
Best Local Similarity 77.6%; Pred. No. 4.01e-03;  
Matches 45; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 1670 tgatattaatgtaatatgaataattaattgtgtgaaaaaaaaaaaaaaaaaaa 1727  
||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1509 TGATTAAATATAATACATATTAATAATGTATGGAGAACCAAAAAAAAAAAAA 1566

RESULT 9  
LOCUS DMU11584 4601 bp DNA INV 23-JUL-1994  
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.  
ACCESSION U11584  
NID 9508826  
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.  
SOURCE fruit fly.  
ORGANISM Mitochondrion Drosophila melanogaster  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;  
Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;  
Epiphytodea; Drosophilidae; Drosophila.  
1 {bases 1 to 4601}  
Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.  
Sequence, Organization and Evolution of the A+T Region of  
Drosophila melanogaster Mitochondrial DNA  
Mol. Biol. Evol. 11, 523-538 (1994)  
MEDLINE 94285822  
REFERENCE 2 (bases 1 to 4601)  
AUTHORS Kaguni,L.S.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI,  
48824-1318, USA  
FEATURES  
source location/Qualifiers  
1..4601  
/organism="Drosophila melanogaster"  
/mitochondrion  
/strain="Oregon-R"  
/dev\_stage="embryo"  
/gene="mt.ori"  
/repeat\_unit 650..1022  
/note="Description: mitochondrial origin; repeat I-A"  
/rpt\_type=tandem

repeat_unit	db_xref="FlyBase:FBgn0013687"	gene="mt:ori"	note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	1025..1360	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	1361..1705	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	1706..2043	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C/A"
repeat_unit	2044..2388	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	2389..2733	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	2734..3078	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	3079..3423	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	3424..3768	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	3769..4113	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	4114..4458	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	4459..4803	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	4804..5148	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	5149..5493	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	5494..5838	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	5839..6183	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	6184..6528	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	6529..6873	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	6874..7218	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	7219..7563	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	7564..7908	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	7909..8253	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	8254..8598	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	8599..8943	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	8944..9288	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	9289..9633	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	9634..9978	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	9979..10323	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	10324..10668	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	10669..11013	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	11014..11358	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	11359..11703	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	11704..12048	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	12049..12393	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	12394..12738	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	12739..13083	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	13084..13428	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	13429..13773	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	13774..14118	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	14119..14463	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	14464..14808	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	14809..15153	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	15154..15498	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	15499..15843	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	15844..16188	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	16189..16533	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	16534..16878	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	16879..17223	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	17224..17568	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	17569..17913	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	17914..18258	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	18259..18603	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	18604..18948	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	18949..19293	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	19294..19638	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	19639..19983	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	19984..20328	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	20329..20673	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B2"
repeat_unit	20674..21018	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	21019..21363	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	21364..21708	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	21709..22053	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B1"
repeat_unit	22054..22398	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	22399..22743	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-C"
repeat_unit	22744..23088	/rpt_type=tandem	/db_xref="FlyBase:FBgn0013687"
repeat_unit	23089..23433	/gene="mt:ori"	/note="Description: mitochondrial origin; repeat I-B





```

RESULT 13      BBGA      1512 bp      RNA      MAM      19-JUL-1996
LOCUS          B. taurus mRNA for cyclin A.
DEFINITION    X68321
ACCESSION     99
NID            cyclin A; protein kinase activation.
SOURCE         domestic cattle or domestic cow.
ORGANISM      Bos taurus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1512)
Hunt, T.
Direct Submission
Submitted (08-SEP-1992) T. Hunt, ICRF Clare Hall Laboratories,
South Mimms, Herts EN6 3LD, UK
REFERENCE     2 (bases 1 to 1512)
AUTHORS      Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J. P., Gannon, J. and
Hunt, T.
TITLE        Identification of the domains in cyclin A required for binding to,
and activation of, p34cdc2 and p32cdk2 protein kinase subunits
MOL. BIOL. CELL 3 (11), 1279-1294 (1992)
JOURNAL       96164440
MEDLINE       93091274
REFERENCE     3 (bases 433 to 1221)
AUTHORS      Brown, N.R., Noble, M.E., Endicott, J.A., Garman, E.F., Wakatsuki, S.,
Mitchell, E., Rasmussen, B., Hunt, T. and Johnson, L.N.
TITLE        The crystal structure of cyclin A.
JOURNAL       Structure 3 (11), 1235-1247 (1995)
MEDLINE       96164440
COMMENT      In the 'Structure' publication: The 5' end of clone A3 was made by
PCR, and introduced a methionine in place of a valine, followed by
a glycine to make an NcoI site. At the 3' end, an XhoI site was
engineered in to allow subcloning into pET21d, which also puts 6
histidines at the C-terminus that were used for purification.
FEATURES      Location/Qualifiers
Source        1..1512
              /organism="Bos taurus"
              /dev_stage="adult"
              /tissue_type="lymphocyte"
              <1..1221
              /note="subunit structure=cyclin A /p34cdc2 or cyclin A
              /p33cdk2"
              /codon_start=1
              /product="cyclin A-3"
              /db_xref="PID:954213"
              /db_xref="PID:g10"
              /db_xref="SWISS-PROT:P30274"
              /translation="EFQDQENVNPEKAAPOQPRFRAGLAVLRAGNSRGAPQRPKT
              RRVAPLKDLPINDEYVPPWPKANNKOPATIHVDEAEIQRPTESKKSSEDLVAF
              NSAVTLQPRKPLAPLDYPMDGSEFSPHTEMVSVLEDEKVPVSYNEVPDYHEDLHTYL
              REMEVCKPKVGMKQPDITNSMRALYDVLVEGEYKQLONETHLAVNYIDRFIS
              SMSVLKQLQVGTAMLLASKEEIYPEVAEFTIIDDITTKQVLRMEHLVRLV
              AFDLAAPTINGLRFQYFLHOOPANCVRVESLAFGLSELDADPFLYKLPVIAAAAF
              HLALYTVGQSWPESLVQKGTGTYLTKPCLLDLHQTYLRAPQHAQOSIREKYKNSKY
              HGVSLNPPETLVN"
BASE COUNT    481 a      298 c      321 g      412 t
ORIGIN
Query Match      1.9%; Score 29; DB 46; Length 1512;
Best Local Similarity 72.3%; Pred. No. 1.57e-01;
Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 1448 tatataatgatgcagctgtccaggataataataataatgatttgataaaaaa 1507
|||||
QY 1496 TATGAATGATGTTGATTAAATATAATACATATTAAATGTATGGAGACCAAAAAA 1555
|||||

Db 1508 aaaa 1512
|||||
QY 1556 AAAAA 1560

RESULT 14      RNP2XMR      1837 bp      RNA      ROD      06-JAN-1995
LOCUS          B. taurus mRNA for cyclin A.
DEFINITION    X68321
ACCESSION     99
NID            cyclin A; protein kinase activation.
SOURCE         domestic cattle or domestic cow.
ORGANISM      Bos taurus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1512)
Hunt, T.
Direct Submission
Submitted (08-SEP-1992) T. Hunt, ICRF Clare Hall Laboratories,
South Mimms, Herts EN6 3LD, UK
REFERENCE     2 (bases 1 to 1512)
AUTHORS      Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J. P., Gannon, J. and
Hunt, T.
TITLE        Identification of the domains in cyclin A required for binding to,
and activation of, p34cdc2 and p32cdk2 protein kinase subunits
MOL. BIOL. CELL 3 (11), 1279-1294 (1992)
JOURNAL       96164440
MEDLINE       93091274
REFERENCE     3 (bases 433 to 1221)
AUTHORS      Brown, N.R., Noble, M.E., Endicott, J.A., Garman, E.F., Wakatsuki, S.,
Mitchell, E., Rasmussen, B., Hunt, T. and Johnson, L.N.
TITLE        The crystal structure of cyclin A.
JOURNAL       Structure 3 (11), 1235-1247 (1995)
MEDLINE       96164440
COMMENT      In the 'Structure' publication: The 5' end of clone A3 was made by
PCR, and introduced a methionine in place of a valine, followed by
a glycine to make an NcoI site. At the 3' end, an XhoI site was
engineered in to allow subcloning into pET21d, which also puts 6
histidines at the C-terminus that were used for purification.
FEATURES      Location/Qualifiers
Source        1..1512
              /organism="Bos taurus"
              /dev_stage="adult"
              /tissue_type="lymphocyte"
              <1..1221
              /note="subunit structure=cyclin A /p34cdc2 or cyclin A
              /p33cdk2"
              /codon_start=1
              /product="cyclin A-3"
              /db_xref="PID:954213"
              /db_xref="PID:g10"
              /db_xref="SWISS-PROT:P30274"
              /translation="EFQDQENVNPEKAAPOQPRFRAGLAVLRAGNSRGAPQRPKT
              RRVAPLKDLPINDEYVPPWPKANNKOPATIHVDEAEIQRPTESKKSSEDLVAF
              NSAVTLQPRKPLAPLDYPMDGSEFSPHTEMVSVLEDEKVPVSYNEVPDYHEDLHTYL
              REMEVCKPKVGMKQPDITNSMRALYDVLVEGEYKQLONETHLAVNYIDRFIS
              SMSVLKQLQVGTAMLLASKEEIYPEVAEFTIIDDITTKQVLRMEHLVRLV
              AFDLAAPTINGLRFQYFLHOOPANCVRVESLAFGLSELDADPFLYKLPVIAAAAF
              HLALYTVGQSWPESLVQKGTGTYLTKPCLLDLHQTYLRAPQHAQOSIREKYKNSKY
              HGVSLNPPETLVN"
BASE COUNT    481 a      298 c      321 g      412 t
ORIGIN
Query Match      1.9%; Score 29; DB 46; Length 1512;
Best Local Similarity 72.3%; Pred. No. 1.57e-01;
Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 1448 tatataatgatgcagctgtccaggataataataataatgatttgataaaaaa 1507
|||||
QY 1496 TATGAATGATGTTGATTAAATATAATACATATTAAATGTATGGAGACCAAAAAA 1555
|||||

Db 1508 aaaa 1512
|||||
QY 1556 AAAAA 1560

RESULT 15      PSNANC      1862 bp      RNA      PLN      07-FEB-1997
LOCUS          Piromyces sp. mRNA for endo-b1,4-mannanase.
DEFINITION    X97520
ACCESSION     9192878
NID            endo-1,4 beta-mannanase; manC gene.
KEYWORDS      Piromyces sp.
SOURCE        Piromyces sp.
ORGANISM      Eukaryotes; mitochondrial eukaryotes; Fungi; Chytridiomycota;
Chytridiomycetes; Neocallimastix; Neocallimastixaceae;
Piromyces
1 (bases 1 to 1862)
Gilbert, H.J.
Direct Submission
Submitted (24-APR-1996) H.J. Gilbert, Department of Biological and
Nutritional Sciences, Univ. of Newcastle upon Tyne, Newcastle, NE1
7RU, UK
REFERENCE     2 (bases 1 to 1862)
AUTHORS      Millward-Sadler, S.J., Hall, J., Black, G.W., Hazlewood, G.P. and

```





[illegible]





KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.
FS	Escherichia coli.
OH	Key
FT	misc_feature 19..69
FT	/*tag= a
FT	/function= multiple cloning site
FT	primer_bind 187..204
FT	/*tag= b
FT	EP-285123-A.
PD	05-MAY-1988.
PF	30-MAR-1988; 105163.
PR	03-APR-1987; US-034819.
PI	(SUSO) SUOMEN SOKERI OY.
PA	Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
PA	WPI: 88-279927/40.
DR	Introducing random point mutations into nucleic acids -
PT	by prep'n of single stranded template, annealing a primer, elongation,
PT	misincorporation, completion of molecules and screening.
PS	Disclosure: P; English.
CC	Random point mutations were introduced into the alpha fragment of
CC	E.coli beta-galactosidase. The wild type sequence was obtained as a
CC	single stranded template and an oligonucleotide was hybridised to
CC	it to generate a pop'n of DNA molecules which terminate at all
CC	possible nucleotide positions within a specified region. The
CC	variable 3' ends generated in this way are used as primers for
CC	reverse transcriptase. Nucleotides are misincorporated by the
CC	transcriptase and the molecules are suitable to forms that can be
CC	amplified and then expressed in a suitable host-vector system.
CC	The sequence covers all 176 diff base substitutions, most of which
CC	occurred singularly in any given mutant.
CC	See also P80575.
CC	Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;
QY	Query Match 2.6%; Score 40; DB 1; Length 204;
	Best Local Similarity 17.0%; Pred. No. 2.56e-07;
	Matches 26; Conservative 63; Mismatches 64; Indels 0; Gaps 0;
Db	39 gycgacbcyrraggnycgcggggywccgagcygcgaayycdchvgcgymrttthhyrrmr 98
	:     :     :     :     :     :     :     :     :     :     :     :
QY	164 GCTGCGGTGTGCGAGCGCGCGCGTCTCAGCGCGCGCTCGGTCTCTCTCTCGGCT 223
	:     :     :     :     :     :     :     :     :     :     :
Db	99 bnvrydrnrsdaawccyrrsvkdcynachhddhyvbbbyvnhnnnccccbnn 158
	:     :     :     :     :     :     :     :     :     :     :
QY	224 GGGAGGGGCGGTATCTCGGGGCGCTGCCAGCCCGCGCGCTCGATATCAAGGGCC 283
	:     :     :     :     :     :     :     :     :     :     :
Db	159 hvchnvbnhrnwayvrhdarrdhvccvchc 191
	:     :     :     :     :     :     :     :     :     :     :
QY	284 TCGGCGGTGCTGCCGACGCTCATTCATCCATCGCC 316
	:     :     :     :     :     :     :     :     :     :     :
RESULT	6
ID	NB1164 standard; DNA; 204 BP.
AC	NB1164;
DT	08-NOV-1990 (first entry)
DE	Base substituted E.coli beta-galactosidase alpha-fragment.
KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS	Escherichia coli.
FT	misc_feature 19..69
FT	/*tag= a
FT	/function= multiple cloning site
FT	primer_bind 187..204
FT	/*tag= b
FT	EP-285123-A.
PD	05-MAY-1988.
PF	30-MAR-1988; 105163.
PR	03-APR-1987; US-034819.
PI	(SUSO) SUOMEN SOKERI OY.
PA	Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
PA	WPI: 88-279927/40.
DR	Introducing random point mutations into nucleic acids -
PT	by prep'n of single stranded template, annealing a primer, elongation,
PT	misincorporation, completion of molecules and screening.









\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 11:10:33 1997; Maspar time 829.51 Seconds  
Tabular output not generated. 969.077 Million cell updates/sec  
Title: >US-08-842-827-3  
Description: (1-1566) from US08842827.seq  
Perfect Score: 1566  
N.A. Sequence: 1 CTTGTGGAGAGAGCGCGG.....CCAAAAA.....GGTTTTTTTTTTTTTT  
Comp: GCACACCTCTCTCGCGGCC.....

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 04  
Listing first 45 summaries

Database:  
EST-STS  
1: EST1 2: EST2 3: EST3 4: EST4 5: EST5 6: EST6 7: EST7 8: EST8  
9: EST9 10: EST10 11: EST11 12: EST12 13: EST13 14: EST14  
15: EST15 16: EST16 17: EST17 18: EST18 19: EST19 20: EST20  
21: EST21 22: EST22 23: EST23 24: EST24 25: EST25 26: EST26  
27: EST27 28: EST28 29: EST29 30: EST30 31: EST31 32: EST32  
33: EST33 34: EST34 35: EST35 36: EST36 37: EST37 38: EST38  
39: EST39 40: EST40 41: EST41 42: EST42 43: EST43 44: EST44  
45: EST45 46: EST46 47: EST47 48: EST48 49: EST49 50: EST50  
51: EST51 52: EST52 53: EST53 54: EST54 55: EST55 56: EST56  
57: EST57 58: EST58 59: EST59 60: EST60 61: EST61 62: EST62  
63: EST63 64: EST64 65: EST65 66: EST66 67: EST67 68: EST68  
69: EST69 70: EST70 71: EST71 72: EST72 73: EST73 74: EST74  
75: EST75 76: EST76 77: EST77 78: EST78 79: EST79 80: EST80  
81: EST81 82: EST82 83: EST83 84: EST84 85: EST85 86: EST86  
87: EST87 88: EST88 89: EST89 90: EST90 91: EST91 92: EST92  
93: EST93 94: EST94 95: EST95 96: EST96 97: EST97 98: EST98  
99: EST99  
EST-STS-TWO  
100: EST100 101: EST101 102: EST102 103: EST103 104: EST104  
105: EST105 106: EST106 107: EST107 108: EST108 109: EST109  
110: EST110 111: EST111 112: EST112 113: EST113 114: EST114  
115: EST115 116: EST116 117: EST117 118: EST118 119: EST119  
120: EST120 121: EST121 122: EST122 123: EST123 124: EST124  
125: EST125 126: EST126 127: EST127 128: EST128 129: EST129  
130: EST130 131: EST131 132: EST132 133: EST133 134: EST134  
135: EST135 136: EST136 137: EST137 138: EST138 139: EST139  
140: EST140 141: EST141 142: EST142 143: EST143 144: EST144  
145: EST145 146: EST146 147: EST147 148: EST148 149: EST149  
150: EST150 151: EST151 152: EST152 153: EST153 154: EST154  
155: EST155 156: EST156 157: EST157 158: EST158 159: EST159  
160: EST160 161: EST161 162: EST162 163: EST163 164: EST164  
165: EST165 166: EST166 167: EST167 168: EST168 169: EST169  
170: EST170 171: EST171 172: EST172 173: EST173 174: EST174

Statistics: Mean 11.818; Variance 2.886; scale 4.094

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
C 1	528	33.7	587 75	H97570	yx03b06.s1	Homo sapie	0.00e+00
C 2	515	32.9	545 137	AA043085	zk48b02.r1	Soares pre	0.00e+00
C 3	507	32.4	552 159	AA036943	zk30f04.s1	Soares pre	0.00e+00
C 4	501	32.0	546 171	W67667	zk38a11.s1	Soares pre	0.00e+00
C 5	461	29.4	465 160	AA040858	zk48b02.s1	Soares pre	0.00e+00
C 6	413	26.4	472 114	W04968	zk43g09.r1	Soares pre	0.00e+00
C 7	414	26.4	490 92	N31047	yx51h07.r1	Homo sapie	0.00e+00
C 8	406	25.9	446 171	W67666	zk38a11.r1	Soares pre	0.00e+00
C 9	405	25.9	456 169	W45106	zk21a10.r1	Soares pre	0.00e+00
C 10	406	25.9	474 113	W01275	zk40f11.r1	Soares pre	0.00e+00
C 11	395	25.2	422 169	W45155	zk21a10.s1	Soares pre	0.00e+00
C 12	381	24.3	456 67	H68363	yr82f11.r1	Homo sapie	0.00e+00
C 13	380	24.3	471 189	AA152123	zk14g02.r1	Soares pre	0.00e+00
C 14	377	24.1	479 74	H90961	yu86e01.r1	Homo sapie	0.00e+00
C 15	376	24.0	415 151	W24605	zb63f08.r1	Soares pre	0.00e+00
C 16	374	23.8	423 78	H97295	Yq74h06.r1	Homo sapie	0.00e+00
C 17	357	22.8	452 85	H54373	Yq93g09.r1	Homo sapie	0.00e+00
C 18	357	22.8	471 85	H57543	Yr05d10.r1	Homo sapie	0.00e+00
C 19	357	22.8	480 169	W45113	zk21c09.r1	Soares pre	0.00e+00
C 20	354	22.6	404 62	H17855	Ym36b12.s1	Homo sapie	0.00e+00
C 21	347	22.2	438 4	T70188	Yc18c09.s1	Homo sapie	0.00e+00
C 22	347	22.2	473 85	H54296	Yq93g09.s1	Homo sapie	0.00e+00
C 23	343	21.9	445 85	H57544	Yr05d10.s1	Homo sapie	0.00e+00
C 24	342	21.8	356 165	C17623	Human placenta	CDNA 5	0.00e+00
C 25	334	21.3	386 5	T70311	Yd26h07.s1	Homo sapie	0.00e+00
C 26	332	20.6	331 158	AA033777	zk19g11.r1	Soares pre	0.00e+00
C 27	321	20.5	351 85	H57213	Yr08d07.r1	Homo sapie	0.00e+00
C 28	319	20.4	332 158	AA033778	zk19g11.s1	Soares pre	0.00e+00
C 29	314	20.1	357 159	AA037575	zk34f08.r1	Soares pre	0.00e+00
C 30	315	20.1	408 7	T80888	Yd36h07.r1	Homo sapie	0.00e+00
C 31	310	19.8	357 64	H26307	Yl31f02.s1	Homo sapie	0.00e+00
C 32	308	19.7	386 159	AA036941	zk30f04.r1	Soares pre	0.00e+00
C 33	304	19.4	413 67	H68364	Yr32f11.s1	Homo sapie	0.00e+00
C 34	298	19.0	422 4	T89858	Yc18c09.r1	Homo sapie	0.00e+00
C 35	287	18.3	304 198	H96367	Yw61f11.s1	Soares pla	0.00e+00
C 36	270	17.2	279 108	H5C37E112	H. sapiens partial	cd	0.00e+00
C 37	269	17.2	315 9	T90107	Yd39g02.s1	Homo sapie	0.00e+00
C 38	265	16.9	287 142	N86994	L1773f fetal heart,	L	0.00e+00
C 39	257	16.4	351 36	R07363	Yc96g08.s1	Homo sapie	0.00e+00
C 40	249	15.9	252 189	AA150023	zk14g01.s1	Soares pre	0.00e+00
C 41	247	15.8	325 128	W92816	Yf80g08.s1	Soares pre	0.00e+00
C 42	248	15.8	331 37	R09946	Yf30f03.r1	Homo sapie	0.00e+00
C 43	237	15.1	247 108	H5C36C022	H. sapiens partial	cd	0.00e+00
C 44	234	14.9	511 154	AA008988	mg99e04.r1	Soares mou	0.00e+00
C 45	226	14.4	354 73	H90307	yu86e01.s1	Homo sapie	0.00e+00

ALIGNMENTS

RESULT 1 H97570 587 bp mRNA EST 12-DEC-1995  
LOCUS Yx03b06.s1 Homo sapiens CDNA clone 260627 3'  
DEFINITION H97570  
ACCESSION g1118455  
NID EST.  
KEYWORDS human clone=260627 primer=ml3 -40 forward library=Soares melanocyte  
SOURCE



2NDHM vector-pT7T3D (Pharmacia) with a modified polylinker host-OHI0B (ampicillin resistant) *SalI*1-Not I *SalI*2-Eco RI Male. 1st strand cDNA was primed with a Not I - Oligo(dT) primer (5'-TGTACCAATCGAATGGAGCCGCGCAGTCTTTTCTTTTCTT-3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. From normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM      Homo sapiens

Deuterostomia; Metazoa; Eumetazoa; Bilateria; Coelomata; Eukaryotae; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chelonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 587)

AUTHORS      Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holmes, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rochford, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE      The WashU-Merck EST Project

JOURNAL      Unpublished (1995)

Contact: Wilson RK  
WashU-Merck Est Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
High quality sequence stops: 351  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

FEATURES	source	Location/Qualifiers	1..587	173 a	117 c	120 g	174 t	3 others
mrna		/organism="Homo sapiens"						
		/clone="280627"						
		<1..>587						
BASE COUNT			173 a	117 c	120 g	174 t		3 others
ORIGIN								
Query Match			33.7%			Score 528; DB 75;	Length 587;	
Best Local Similarity			96.6%			pred. No. 0.00e+00;		
Matches	565;	Conservative	0;	Mismatches	16;	Indels	4;	Gaps 4;

Db	1	ggtctcccatcaacttttaataatgattattattatcaaatcaaacatcatctcatagaaagca	60
Cp	1548	GGTTCTCCCATACATATTTAAATGATGTATATATTAAATCAACATCATTCATAGAAAGCA	1489
Db	61	tattacatacatgctttatatacaataagcattacactttttttaaataaaatgatacacaggagg	120
Cp	1488	TATTACATACATGTTTATACATAAAGCATTACATTTTTTTTTTAAATAAATGTATACAGGTGG	1429
Db	121	ggcactgttttggtagaaggcttggaggttttttaataagcttttagagctattagataaac	180
Cp	1428	GGCAGCTGTTTTGGTGGAAAGCTTGGAGTTTTTTTAAATGAGTTAGAGCTATTAGATAACC	1369
Db	181	actgagtttaaaaggttaactatgtacacacaaagtg-gatccaagaaggcatacagaagca	239
Cp	1368	ACTGAGTTAAAGGTAACTATGTACACACAAAGGTGTGCATCCAGAGAGCATATACAGACGA	1309
Db	240	gaagctctttaaaggcttgtgtaacacagggaagaagatgatactctctgccttggtaaatac	299
Cp	1308	GAAGTCTTTAAAGCTTGTAACACAGGAAGAAAGATGATCTCTTGGCTTGTGGCAATC	1249
Db	300	attttccctttagaaacaggccagcttcaacctggggaccacctgtgcctttcaaggctggg	359
Cp	1248	ATTTTCCCTTTAGAAACAGGCCAGCTTCACTTGGGACACCTGTGTGCTCTTCAAGGCTGG-	1190
Db	360	tgattgctcggaatagttatcccaagttgttggtgtttcatgcagagttgtgatgagatgcc	419

Cp	1189	TGATTGCTCGGATAGTATCCAGATTGTTGGTGTTTCATGTCAGAGATTGTATGAGAGTCC	1130			
Db	420	tctctcttctctttaaagaagctcttctcccttgagaanaatccgatacatatacagca	479			
Cp	1129	TGCTCTTTTCTTTCTTTTAAAGAAGTCTTTCTTTTGAAGAAATCCGATACATATACGCA	1070			
Db	480	actaatattgcaacagagctccctgaatgagtcagctcaacacatcgctcccaagtggg	539			
Cp	1069	ACTATATTTGCACACAGAGCTCCCTGTAATGAGTCCAGTCAACACATCGGTCC	1011			
Db	540	gtttataatccgnaactccgngaaggncccaataaatgggaagg	584			
Cp	1010	-TTTATAATCAGAACTCGAAGAAGGCCACATAAATGGGATACGG	957			
RESULT	2	AA043085	545 bp	mRNA	EST	04-SEP-1996
LOCUS		2k48b02.r1	Soares pregnant uterus	NbHPU	Homo sapiens	CDNA clone
DEFINITION		48027 5				
ACCESSION		AA043085				
NID		91520979				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS		1 (bases 1 to 545)				
		Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hulman,M., Kudaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasak,E., Waterston,R., Williamson,A., Wohldmann,P. and Willson,R.				
TITLE		The WashU-Merck EST project				
JOURNAL		Unpublished (1995)				

Contact: Wilson RK  
 Washo-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estawatson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28M13 rev2 from AmerSham  
 High quality sequence stop: 485.  
 Location/Qualifiers  
 1..545  
 /organism="Homo sapiens"  
 /note="Organ: uterus; Vector: pTV73-Pac; Site\_1: Not I;  
 Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dN) primer [5',  
 AACTGGAGAGATTGCGGCCCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 and Eco RI sites of the modified pTV73 vector. Library  
 went through one round of normalization. Library  
 constructed by M. Fatima Bonalido."

```

clone="486027" /clone_lib="Soares pregnant uterus NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DHIOB"
mrna <1..>545
BASE COUNT      159 a   116 c   105 g   163 t       2 others
ORIGIN
Query Match          32.9%; Score 515; DB 137; Length 545;
Best Local Similarity 98.7%; Pred. NO. 0.00e+00;
Matches 540; Conservative 0; Mismatches 3; Indels 4; Gaps 4;
Db 1 ccacactgcaatttgcttggcgtatccattatgtggcgcttccttcgmgtttct 60
|||||

```





REFERENCE 1 (bases 1 to 546)  
 AUTHORS Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 Hillier, B., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Wilton RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 634 Std Error: 0.00  
 Seq primer: mob.REGA-ET  
 High quality sequence stop: 321.  
 FEATURES  
 source  
 1..546  
 /organism="Homo sapiens"  
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCACTCAAGTGGAGCGCGCATCTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M.Fatima Bonaudo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NBHL19W."  
 /clone\_lib="Soares fetal heart NBHL19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 complement(<1..>546)  
 BASE COUNT 159 a 105 c 121 g 159 t 2 others  
 ORIGIN  
 Query Match 32.0%; Score 501; DB 171; Length 546;  
 Best Local Similarity 97.6%; Pred. No. 0.00e+00;  
 Matches 523; Conservative 0; Mismatches 11; Indels 2; Gaps 2;  
 Db 2 ttatatacagaattccatttttttaataaaaaatgtatcacagggtgggacctgttttg 61  
 Cp 1475 TTTATACATAAGCATTACATTTTTTTTAATAAAAAATGTATACAGGTGGGCGACTGTGTGG 1416  
 Db 62 tgaaggcttgagttttttaaagtattagactattagataaccactgactgattaaagg 121  
 Cp 1415 TGGAGGCTTGGAGCTTTTATAGTTAGTTAGCTATTAGATACCACTGATGTTAAGG 1356  
 Db 122 taactatatacacacaagaatggcnatccaggagcagatagacagacagaagtctttaag 181  
 Cp 1395 TAACATATGTACACACAAGGTGCATCCAAAGAGGCATATACAGCAGCAAGAATCTTTAAG 1296  
 Db 182 gcttgatcacaggagaagaagatgcatcctcttgctgttg9-aatcatctttctcttaga 240  
 Cp 1295 GCTTGATACACAGGAAGAAGATGATCCCTCTTGCCTTGTGGCAATCATTTTCTTTAGA 1236  
 Db 241 aaacagcgcttcaactctgggacacctgtgcctttcaaggctggtatgctcggata 300  
 Cp 1235 AAACAGGCGACCTTCACTCGGACACCCCTGCTCTTCAAGCTGGTATGTCGCGATA 1176  
 Db 301 gtgattccagttgttggttttcaatgcagagttgtatgagagtcctcctcttttttc 360  
 Cp 1175 GTGATTCACAGTTGTGGTGTTCATGCAGAGTTGTATGAGAGTCCTCCTCTTCTTTC 1116

Db 361 tttaaaagaagttcttcttttgagaataatccgatacatatcacagaactaatattgcaac 420  
 Cp 1115 TTTAAAAGAAGTTCTTCTTTGAGAAATCCGATACATATACAGCAACTAATATTGCAAC 1056  
 Db 421 cagagctccctgaagtagctccagtcacacatcgctccagtcagtggtttataatcagaaa 480  
 Cp 1055 CAGAGCTCCCTG-AATGAGTCCAGTCAACACATCGTCCAGTGGTGTATTATATCAGAAA 997  
 Db 481 ccccgagaaaggccacataaattggtacggcaacaagaccacaaattgcagtgtggg 536  
 Cp 996 CTCGAGAAGGCCACATAAATGGATACGGCAACAAGACCAAAATTCAGTGTGGG 941  
 RESULT 5  
 LOCUS AA040858 465 bp mRNA EST 30-AUG-1996  
 DEFINITION ZK48502.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone  
 486027 3'  
 ACCESSION AA040858  
 NID g1517154  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 465)  
 AUTHORS Hillier, B., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Wilton RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40M13 fwd. from Amersham  
 High quality sequence stop: 354.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..465  
 /organism="Homo sapiens"  
 /note="Organ: uterus; Vector: pT73-Pac; Site\_1: Not I;  
 Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5',  
 AACTGGAAGATTCGGCGCGCTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by M. Fatima Bonaudo."  
 /clone="486027"  
 /clone\_lib="Soares pregnant uterus NbHPU"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 complement(<1..>465)  
 BASE COUNT 134 a 83 c 93 g 153 t 2 others  
 ORIGIN

Query Match 29.4%; Score 461; DB 160; Length 465;  
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;  
 Matches 462; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 1 ggtctccatcacatttattatgtattattataatcaaacatcattcatgagaagca 60  
 Cp 1548 GGTCTCCCATACATTTAATATGTTATATATTTAAATCAACATCATTCATTAGAAAGCA 1489



Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 385  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
Source 1..490  
/organism="Homo sapiens"  
/clone="265309"  
<1..>490

BASE COUNT 131 a 107 c 111 g 138 t 3 others

ORIGIN  
Query Match 25.4%; Score 414; DB 92; Length 490;  
Best Local Similarity 97.1%; Pred. No. 0.00e+00;  
Matches 476; Conservative 0; Mismatches 4; Indels 10; Gaps 10;

Db 1 gngcctcactcttggatgttggatcgatcgatggtcacaatacaactcagcgtg 60  
|||  
QY 722 GCGGCGCTCACTTCCTGGATGTTTGTATCGATCGATGGTCAAAAATCAACTGCGCGATGG 781  
|||  
Db 61 ttacatgaactacacatagtcgaggaatcagaaanagttaaggaagcaggtgtc 120  
|||  
QY 782 TTACATTTGAATACATATGTCGAGGGAATGCGAAGAGTTAAGGAGGAGGTTGTC 841  
|||  
Db 121 ttctatcggcgcactcttcgttttccatgtactgcgtgctgttggcactttatct 180  
|||  
QY 842 CTTCTATTCAAGGCGACCTCTCGTGTTCCTCATGCTGATCGATGCTGTTTGGCATTATCT 901  
|||  
Db 181 tcaagccaggtgagggagactggcagactcttcagcccccacactgcaattgtct 240  
|||  
QY 902 TCAAGCCAGGATGAAGGAGACTTGGCAGACTCTTACGCCACACTGCAATTTGGTCT 961  
|||  
Db 241 tgttcgcgtatccatttatgtgggccttcttcgagttcttgattataaacaccactggag 300  
|||  
QY 962 TGTTGCCGTATCCATTTATGTGGGCGCTTCTCGAGTTTCTGATTATAAACACCACCTGGAG 1021  
|||  
Db 301 cgaatgtgtgactgagactcattcaggagactctgttgcaatattagttgctgtatgt 360  
|||  
QY 1022 CGATGTTGACTGGACTCAATTCAGGAGCTCTGGTTGCAATATTAGTTGCTGATATGT 1081  
|||  
Db 361 atcgatttcttcaaaagaagaactctctttaaagaagaagaaggagactctcca 420  
|||  
QY 1082 ATCGGATTTCTTCAAGAAAGAACTTC-TTTTAAAGAAAGAAAG-AGGAGGACTCTC-A 1138  
|||  
Db 421 tacaactcctgcatgaaacacacaaactcgggactcactaaatccggagcaatccacc 480  
|||  
QY 1139 TACAACCTC-TGC-ATG-AAACACCACTGGGAAT-CACTA-TCCG-AGCAATC-ACC 1191  
|||  
Db 481 agccttgaag 490  
|||  
QY 1192 AGCCTTGAAG 1201  
|||

RESULT 8 W67666 446 bp mRNA EST 15-OCT-1996  
LOCUS 2438a11.r1 Soares fetal heart NBHH19W Homo sapiens cDNA clone  
DEFINITION 342908 5'.  
ACCESSION W67666  
NID 91376761  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 446)  
AUTHORS Hillier, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 634 Std Error: 0.00  
Seq Primer: mob.REGA-EF  
High quality sequence stop: 328.  
Location/Qualifiers  
1..446  
/organism="Homo sapiens"  
/note="Organ: heart; Vector: pT773D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGGAGCGCGCATCTTTTCTTTTCTTTT 3'] double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Patima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHH19W."  
/clone="342908"  
/clone\_lib="Soares fetal heart NBHH19W"  
/sex="unknown"  
/lab\_host="DH10B (ampicillin resistant)"  
BASE COUNT 122 a 102 c 100 g 121 t 1 others  
ORIGIN  
Query Match 25.9%; Score 406; DB 171; Length 446;  
Best Local Similarity 98.9%; Pred. No. 0.00e+00;  
Matches 431; Conservative 0; Mismatches 1; Indels 4; Gaps 4;  
Db 2 ttatcttcaagccaggtgagggagactggcagactcttcagcccccacactgcaat 61  
|||  
QY 895 TTTATCTTCAAGCCAGGATGAAGGAGACTGGGCAAGACTCTTACGCCCCACACTGCAAT 954  
|||  
Db 62 ttggccttgttgcgtatccatttatgtgggccttcttcgagttcttgattataaacacc 121  
|||  
QY 955 TTGCTCTGTGTGGCGTATCCATTTATGTGGGCGCTTTCGAGTTTCTGATTATAACACC 1014  
|||  
Db 122 actggagcgtatgttggactggactcattcaggagactcctgttgcaattattagttgctg 181  
|||  
QY 1015 ACTGGACCGATGTGTGACTGGACTCATTCAGGGAGGCTCTGTTGCAATATTAGTTGCTG 1074  
|||  
Db 182 tatatgtatggatttcttcaagaagaagaactctctttaaagaagaagaaggagagact 241  
|||  
QY 1075 TATATGTATCGGATTTCTTCAAGAAAGAACTTCCTTTTAAAGAAAGAAAGAGGAGGACT 1134  
|||  
Db 242 ctacatacaactctgcgtgaaacacacaaactcgggaatcactatcccgagcattcaccagc 301  
|||  
QY 1135 CTCATACAACCTCTGCATGAACACCAACCACTGGGAATCACTATCCGAGCAATCACCAGC 1194  
|||  
Db 302 cttgaaggcagcaggggtgcc-agggtgaagctggcctgttttctaaaggaaaaatgattgc 360  
|||

```

QY 1195 CTTGAAGGCAGCAGGGTCCAGGTGAAGTGGCTGTTTCTTAAGGAAATGATTC 1254
Db 361 cacaaggcaagaagatgcatcttcttctctggtgtacaaagcctttaaagactctggc 420
QY 1255 CACAAGGCAA-GA-GGATGCATCTTCTCTCTGTTGTGTACAAAGCCTTAAAGACTTCG-C 1311
Db 421 tgcgtgatgcctctt 436
QY 1312 TGTGTATGCTCTT 1327

RESULT 9
LOCUS W45106 456 bp mRNA EST 10-OCT-1996
DEFINITION zc21a10.r1 Soares senescent fibroblasts NBHSF Homo sapiens cDNA
clone 322938 5'.
ACCESSION W45106
NID g1329189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLM: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 728 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 435.
Location/Qualifiers
1..456
/organism="Homo sapiens"
/note="Vector: p7T3D (Pharmacia) with a modified
polylinker V-type; phagemid. Site.1: Not I; Site.2: Eco
RI; TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTTCTTTT
3' ], double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified p7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/clone="322938"
/clone_lib="Soares senescent fibroblasts NBHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
<1..>456

BASE COUNT 124 a 98 c 107 g 126 t 1 others
ORIGIN

Query Match 25.9%; Score 405; DB 169; Length 456;
Best Local Similarity 98.6%; Pred. No. 0.00e+00;
Matches 436; Conservative 0; Mismatches 1; Indels 5; Gaps 4;

Db 15 ctttatctcaagcaggatgaaggagactgggcaagactcttacgcccccaactgcga 74
QY 894 CTTTATCTTCAAGCCAGGATGAAGGAGACTGGCAAGACTCTTACGCCACACTGCAA 953
Db 75 ttggctctgtccgtatccatttatgtggcccttctctcagttctgtattataaacac 134

```

```

QY 954 TTGTGCTTGTGGCGTATCATTTATGTGGCTTCTCGAGTTCTGATTATTAACAC 1013
Db 135 cactggagcgtgtgtgactgactcattcaggagcgtctgtgttcaataattgtgt 194
QY 1014 CACTGGAGCGATGTCTTGCATCGGACTCATTCAGGAGCTCTGTTGCAATATTAGTTGT 1073
Db 195 gtatatgtcgtattcttcaagaagaagactcttttaagaagaagaagagagagac 254
QY 1074 GTATATGTATCGGATTCTTCAAGAAGAAGAACTCTTTTAAAGAAGAAGAGAGGAC 1133
Db 255 tctcatacaactctgcatgaacacccaactgggaatcactatcccgagcaatcaccag 314
QY 1134 TCTCATCAAACTCTCATGAACACCAACCACTGGGAATCCTATCCGAGCAATCACCAG 1193
Db 315 ccttgaagcagcaggtgcccagctgaagctgcctgttttctaaagaaagaaatgat 374
QY 1194 CTTTGAAG-GGCAGCGGGTCCCGGTGAAGCTGGCCT-GTTTTCTTAAGGAATGAT 1251
Db 375 tgccacaaggcaaggatgcattcttctcctgggtgtgacaagccttttaagacttc 434
QY 1252 TGCACAAGGCAAGAGGATGCATCTTCTTCTCTGG--TGTACAAGCCTTT-AAAGACTTC 1308
Db 435 tgctactctatgcctcttga 456
QY 1309 TGTCTGTATATGCTCTTGA 1330

RESULT 10
LOCUS W01275 474 bp mRNA EST 18-APR-1996
DEFINITION za40f11.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
295053 5' similar to contains element MER12 repetitive element ;.
ACCESSION W01275
NID 91273255
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLM: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: Eppriemer
High quality sequence stop: 429.
Location/Qualifiers
1..474
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I oligo(dT) primer
15' AACGGGAAGTAATTAAGAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
/clone="295053"
/clone_lib="Soares fetal liver spleen INFLS"

```

```

/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
<1..>474
BASE COUNT      133 a   96 c   101 g   138 t   6 others
ORIGIN

Query Match      25.9%; Score 406; DB 113; Length 474;
Best Local Similarity 96.7%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 6; Indels 10; Gaps 10;

Db 1 gagactggcgaactcttaagcccccacacgcgaattggtttgttccgtatccattt 60
QY 919 GAGACTGGCGAAGACTCTTACGCCCCACACATGCAATTTGGCTGTTCGCTATCCATTT 978

Db 61 atgtgggcttctcgagttctgattataacacacacgagcagcatgtttgactggac 120
QY 979 ATGTGGGCTTCTCGAGTTCTGATTATAAACACACCTGGAGCGGATGTGTGCTGGAC 1038

Db 121 tcaattcaggagctctgttgcgaattattgctgtctgtatctgtatctgcgattcttcaaag 180
QY 1039 TCATTCAGGAGCTCTGGTTGCAATATTAGTTGCTGTATATGTATCGGATTTCTTCAAG 1098

Db 181 aaagaactcttttaagaagaagaaagagagagcactctcatacaactctgcataaacac 240
QY 1099 AAGAAGACTCTTTTAAGAAGAAGAAAGAGGAGGACTCTATCAACTCTGCATGAACAC 1158

Db 241 caacaactgggaatcaactatccagcaatcaccagccttgaaggcagcagggtgccag 300
QY 1159 CAACAACCTGGGAATCACTATCCGAGCAATCACCAGCCTTGAAGGCGAGCGGTGCCAG 1218

Db 301 gtgaagctggcctgtttctaaaggaaatgattgcacaaag-aagaggaatgatcttt 359
QY 1219 GTGAAGCTGGCCTGTTTCTAAGAAATGATTGCCAAGCAAGGAGGATGTCATCTTT 1278

Db 360 ctctcgtggtgtn-aagcctttaagactn-tgctgtgt-tatnc-tcttgatgcacact 415
QY 1279 CTCTCGTGTGACAGGCTTTAAGAGCTTCTGCTGTGTATGCTCTTGGATGCACACT 1338

Db 416 ttgtgtgtacatagttac-tttan-tcagtggtta-ctaatan-tctaaa-tcataaaaa 470
QY 1339 TTGTGTGTACATAGTTACTTACTCTACGCTGGTGTATCTAATAGCTCTAAACTCATTA 1398

Db 471 aa 472
QY 1399 AA 1400

RESULT 11
LOCUS      W45155      422 bp      mRNA      EST      10-OCT-1996
DEFINITION Zc21a10.s1 Soares senescent fibroblasts NBHSF Homo sapiens cDNA
            clone 32938 3'.
ACCESSION  W45155
NID         91329246
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 422)
AUTHORS   Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
            Wilson, R.
TITLE      The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel.: 314 286 1800

/organism="Homo sapiens"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
RI; TGTTCACCAATCTGAAGTGGGCGCGCATTTTITTTTTTTTTT
3'), double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/clone_lib="Soares senescent fibroblasts NBHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>422)
BASE COUNT      124 a   75 c   90 g   133 t   2 others
ORIGIN

Query Match      25.2%; Score 395; DB 169; Length 422;
Best Local Similarity 98.3%; Pred. No. 0.00e+00;
Matches 415; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

Db 1 ggtctccatcacattttaatatgtattattataataaacaacatcattctatagaagca 60
Cp 1548 GGTCTCCCATACATTTTAATATGTATTATTTAAATCAAAACATCATATCATAAAGCA 1489

Db 61 tattacatacatgttttatcacataagcattacatttttttaataaaaaatgtatacagtg 120
Cp 1488 TATTACATACATGTTTATACATAAGCATTTACATTTTTTTTAAATAAAATGTATACAGTGG 1429

Db 121 ggcactttttgtggaagccttgaggatttttttaagttagagctattagataaac 180
Cp 1428 GGCACCTGTTTGGTGGGAAGCCTTGGAGTTTTTTTAAATGAGTTTAGAGCTATTAGATAACC 1369

Db 181 actgagttaaaggtaactatgtacacacaaagtgtgcacccaaggtcccaagggcagcagca 240
Cp 1368 ACTGAGTTAAAGGTTAACTATGTACACACAAAGTGTGCATCCAAAGAGGCATATCAGCAGCA 1309

Db 241 gaagctttaaaggctgtgtacacacaggaagaagatgcctctctgcttctggaatc 300
Cp 1308 GAAGCTTTAAAGCTTGTACACACGAAGAAGAATGATCTCTTGTGCTTGTGGCAATC 1249

Db 301 atttccnttagaaaacagccagcttcacctgggacccctgctgccttttcaaggctgg 360
Cp 1248 ATTTTCCTTTAGAAAACAGGCCAGCTTCACCTGGGCACCCCTGCTGCCTTT-CAAGGCTGG 1190

Db 361 tggattctcgagatagtggtcccgattgttggngtttttcagagagttatagagagt 420
Cp 1189 TG-ATTGCTCGGATAGTG-ATTCCTCAGTTGTGGTGTTTCATGACAGAGTTGTATGAGAGT 1132

Db 421 cc 422
Cp 1131 CC 1130

RESULT 12
LOCUS      H58363      456 bp      mRNA      EST      18-OCT-1995
DEFINITION Y182f11.r1 Homo sapiens cDNA clone 211821 5'.
ACCESSION  H58363
NID         91027103
KEYWORDS   EST.
SOURCE     human clone=211821 primer=M13rp1 library=Soares fetal liver spleen
            INFLS vector=pT73D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver

```



QY	1061	AAATATTAGTTCGTATATGTATCGGATTCTTCAAGAAAGAACTTCTTTTAAAGAAAG	1120
Db	247	aaaagaggaggactctcatcaactctgcatgaacacaaactgggaatcaactatcc	306
QY	1121	AAAAGAGGAGGACTCTCATACAACTCTGCATGAACACCAACAACCTGGGAATCACTATCC	1180
Db	307	gagcaatcaccagcccttgaagcgcaagcagggtgcccaagtgtaatctgagccctgttttc	366
QY	1181	GAGCAATCACCAGCCTTGAAGGSCA-GCAGGGTGCCCA-GGTGAA-GCTGGCCTGTTTTC	1237
Db	367	tataaganaatgatgtccacaaaggcaaggagatgcattcttcttcctgggtgtacaag	426
QY	1238	TAAAGHAAATGATGCCACAGGCAAGAGG-ATGCACTTTCTT-CCTGG-TGTACAAG	1294
Db	427	cccttttaagacactctgctgctgcatgcctcttga	465
QY	1295	CC---TTTAAAGAC-TTCTGCTGCTGATATGCTCTTGA	1330

RESULT	14
LOCUS	H90961
DEFINITION	479 bp mRNA EST
ACCESSION	YU86601.r1 Homo sapiens cDNA clone 240696 5'.
NID	H90961
KEYWORDS	G1081391
SOURCE	E5T.
	human clone=240696 primer=M13RP1 library=Soares fetal liver spleen INFLS vector=pr7r3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Raitel-Pac I Reite2-Eco RI Liver and spleen from a 20 week post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', CACTGGAGAGATTAATTAAAGAACCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pr7r3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.  Homo sapiens  Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.  1 (bases 1 to 479)  Hollan,L., Clark,N., Kuebaque,T., Elliston,K., Hawkins,M., Hillman,M., Hultman,M., Kubacka,T., Le,M., Lennon,G., Maria,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  The WashU-Merck EST Project Unpublished (1995)
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

```

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 364
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
    source
        1..479
        /organism="Homo sapiens"
        /clone="240696"
        <1..>479
BASE COUNT      131 a      99 c      102 g      142 t      5 others
ORIGIN
Query Match      24.1%; Score 377; DB 74; Length 479;
Best Local Similarity 97.9%; Pred. No. 0.00e+00;
Matches 420; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

```

Db	1	atttgctcngttgcgcatccattatgtggtggccttctcagagttctcgtatataaaca	60
Qy	953	ATTTGGCTCTGTGGCGGTATCCATTTATGTGGCCTTCTCGAGTTTCTGATTATAACA	1012
Db	61	ccactggagcgatgtgttgactggaacatccaggagcctcgtgtgcaattagttgc	120
Qy	1013	CCACTGGAGCGATGTGTGACTGGACTCATTCAGGAGCTCTGGTTGCAATATTAGTTGC	1072
Db	121	tgatatgtatcggtattcttcttaagaagaactcttcttaagaagaagaagagga	180
Qy	1073	TGTATATGTATCGGATTTCTTCAAGAAAGAACTCTCTTTAAAGAAAGAAAGAGGAGGA	1132
Db	181	ctctcacaactctgcatgaaacaccacaactgggaatcactatccgagcaatcacca	240
Qy	1133	CTCTCATCAACTCTGCATGAACACCAACACTGGGAATCTACTATCCGAGCAATCACCA	1192
Db	241	gccttgaaggcagcaggggtgccccaggtgaagctgctgtttcttaaggagaaatgatt	300
Qy	1193	GCCTTGAAGGCGAGCAGGGTGCCAGAGTGGAAGTGCCCTGTTTTCTTAAAGGAAATGATT	1252
Db	301	gccacaaggcagggagtgatcttctctcctgggtgtacaaagcctttaaagactctcg	360
Qy	1253	GCCACAAGGCAAG-AGGATGCATCTTCTTCTCTGG-TGTACAGCCTTTAAAGACTTCTG	1310
Db	361	ctgctctatgctcttgggattggcacacttctgtgtgtacatagtttaccctttaaactc	420
Qy	1311	CTGCTGATATGCTCTTGG-ATG-CACACTTTGTGTGT-ACATAGTT-ACCTTT-AACTC	1365
Db	421	agtgggtat 429	
Qy	1366	AGTGGTTAT 1374	
RESULT	15		
LOCUS	W24605	415 bp	mRNA
DEFINITION	zb63f08.r1 Soares fetal lung NBHL19W Homo sapiens CDNA clone 308295	5'	EST 20-AUG-1996
ACCESSION	W24605		
NID	g1301591		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;		
AUTHORS	Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,		
	Holman, M., Rulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,		
	Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,		
	Waterson, R., Williamson, A., Wohlmann, P. and Wilson, R.		
TITLE	WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK		
	WashU-Merck EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LLNL ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert Length: 773 Std Error: 0.00		
	Seq primer: mob.REGA+ET.		
FEATURES	Location/Qualifiers		
source	1..415		
	/organism="Homo sapiens"		
	/note="Organ: lung; Vector: p7T73D (Pharmacia) with a		
	modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer		
	[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTTTTT-3']		
	double-stranded cDNA was size selected, ligated to Eco RI		
	adapters (Pharmacia), digested with Not I and cloned into		
	the Not I and Eco RI sites of a modified p7T73 vector		





**This Page Blank (uspto)**

\*\*\*\*\*

MAPSRELH

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 11:27:18 1997; Maspar time 421.77 Seconds  
Tabular output not generated. 1020.354 Million cell updates/sec

Title: >US-08-842-827-3  
Description: (J-1566) from US08842827.seq  
Perfect Score: 1566  
N.A. Sequence: 1 CGCTGGGAGAGACGCGCG.....CCAAAAA.....GGTTTTTTTTTTTTTT  
Comp: GGACACCTCTCTCGCGGC.....

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

EST-STS-THREE

1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13

EST-STS-FOUR

49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnEST5 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnESTs 107:ueBST1  
108:ueBST2

Statistics: Mean 11.842; Variance 3.114; scale 3.803

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	482	30.8	514	28	AA024599	ze78b09.r1 Soares fet	0.00e+00
2	380	24.3	390	9	AA132640	zo20d08.r1 Stratagene	0.00e+00
3	376	24.0	415	87	HS605336	zb63f08.r1 Soares fet	0.00e+00
4	359	22.9	376	9	AA132534	zo20d08.r1 Stratagene	0.00e+00
5	352	22.5	378	28	AA024600	ze78b09.r1 Soares fet	0.00e+00
6	345	22.0	377	30	AA062861	zf70e10.r1 Soares pin	0.00e+00
7	346	22.0	377	30	AA062861	zf70d09.r1 Soares pin	0.00e+00
8	286	18.3	312	30	AA062853	ml95b10.r1 Stratagene	0.00e+00
9	259	16.5	385	33	AA107326	ml95b10.r1 Stratagene	0.00e+00
10	229	14.6	542	22	AA210789	zr30e08.r1 Soares NBH	4.47e-238
11	229	14.6	542	54	AA210789	zr30e08.r1 Soares NBH	4.47e-238
12	120	7.7	241	87	HS368337	zc68b11.r1 Soares fet	2.15e-130
13	119	7.6	501	43	G23282	human STS W15758	6.68e-129
14	88	5.6	313	91	MM1155742	mx78g01.r1 Soares mou	1.86e-83
15	88	5.6	313	66	AA237956	mx78g01.r1 Soares mou	1.86e-83
16	76	4.9	564	88	HS942332	zc64f09.r1 Soares fet	1.97e-66
17	61	3.9	564	1	W30942	zc64f09.r1 Soares fet	1.97e-66
18	58	3.7	411	50	AA152779	mr78c03.r1 Stratagene	7.62e-46
19	53	3.4	87	33	AA106892	ml85a07.r1 Stratagene	7.54e-42
20	53	3.4	242	27	W39815	308 Mouse VM CDNA lib	2.64e-35
21	53	3.4	337	27	W39811	307 Mouse VM CDNA lib	2.64e-35
22	46	2.9	320	10	AA015411	mh21g01.r1 Soares mou	1.96e-26
23	44	2.8	478	26	N75714	mv31b01.r1 Soares fet	5.64e-24
24	42	2.7	185	90	MMG272	ma59e06.r1 Soares mou	1.49e-21
25	40	2.6	184	95	MMG3211	mb78e01.r1 Soares mou	3.55e-19
26	40	2.6	199	10	AA014140	mb78e01.r1 Soares mou	3.55e-19
27	40	2.6	269	95	MM5326	mb22b07.r1 Soares mou	3.55e-19
28	37	2.4	52	49	AA122946	mq69e06.r1 Stratagene	1.07e-15
29	37	2.4	52	6	AA122946	mq69e06.r1 Stratagene	1.07e-15
30	34	2.2	369	24	AA217401	mu30f02.r1 Soares mou	2.42e-12
31	34	2.2	429	56	AA144221	mr14h12.r1 Soares mou	2.42e-12
32	33	2.1	236	81	HS1146241	zp52f10.r1 Stratagene	2.95e-11
33	33	2.1	236	12	AA182434	zp52f10.r1 Stratagene	2.95e-11
34	33	2.1	301	56	AA138853	mr03a06.r1 Soares mou	2.95e-11
35	29	1.9	181	67	AA241676	MBL2SJBST3 JH096SL-B	4.34e-07
36	29	1.9	181	79	BM4A1676	MBL2SJBST3 JH096SL-B	4.34e-07
37	29	1.9	197	70	AA208192	mv91b08.r1 GuayWoodfo	4.34e-07
38	29	1.9	197	20	AA208192	mv91b08.r1 GuayWoodfo	4.34e-07
39	29	1.9	197	103	MAA81921	mv91b08.r1 GuayWoodfo	4.34e-07
40	29	1.9	339	1	N97773	1238C3 czappPDg2 1. D	4.34e-07
41	30	1.9	446	12	AA181403	zp52f10.r1 Stratagene	4.22e-08
42	28	1.8	153	15	AA189532	mt90c05.r1 Soares mou	4.25e-06
43	28	1.8	244	5	AA117061	mn29d08.r1 Heddington	4.25e-06
44	28	1.8	453	105	MMW9541	mf84g07.r1 Soares mou	4.25e-06
45	28	1.8	1097	73	C22771	Dictyostellium discoid	4.25e-06

ALIGNMENTS

RESULT	1	AA024599	514 bp	mRNA	EST	01-FEB-1997
LOCUS		ze78b09.r1	Soares fetal heart NBH19W	Homo sapiens	CDNA clone	
DEFINITION		365081 5'				
ACCESSION		AA024599				
NID		q1489504				
KEYWORDS		EST				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrate; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS		1 (bases 1 to 514)				
		Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, P., Trevaskis, E., Watson, R., Williamson, A., Wohlmann, P. and Wilson, R.				
TITLE		WashU-Merck EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine				

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 600 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 478.

# FEATURES

Source

1..514

/organism="Homo sapiens"

/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGGAGCGCGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Patima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."

/clone="365081"

/clone.lib="Soares fetal heart NBHL19W"

/sex="unknown"

/dev\_stage="19 weeks"

/lab\_host="DH10B (ampicillin resistant)"

<1..>514

BASE COUNT 162 a 98 c 97 g 153 t 4 others

# ORIGIN

Query Match 30.8%; Score 482; DB 28; Length 514;  
Best Local Similarity 98.4%; Pred. No. 0.00e+00;  
Matches 507; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Db 1 ctggactcattcaggagctcgtgtgcaatattagttgctgtatatgtatcggaattctt 60  
QY 1033 CTGGACTCATTCAGGAGCTCTGTTGCAATATTAGTTGCTGTATATGTCGGATTCTT 1092

Db 61 tcaagaagaactcttttaagaaganaagaagagactctatacaactctgcgatg 120  
QY 1093 TCAAGAAGAAGACTCTCTTTAAGAAGAAGAAGAGGAGGACTCTCATCAACTCTGCAATG 1152

Db 121 aaacaccacaactgggaatcactatccgagcaatcaccagccttgaaggcagcagggt 180  
QY 1153 AAACACCAACAACTGGGAATCACTATCCGAGCAATCACCAGCCTTGAAGGCGAGCGGT 1212

Db 181 gccacgggtgaagctggcctgtttctaaaggaatgattgccaaagn-aagaggatgc 239  
QY 1213 GCCAGGTGAAGCTGGCTGTCTTAAAGAAAATGATTGCCAAGCAAGAGGATGC 1272

Db 240 atctttctcctgtgtacaaagcctttaaagactctctgtctgtatgcctcttggatg 299  
QY 1273 ATCTTTCTCTGCTGTGTACAGCCTTTAAAGACTTCTGCTGTATATGCTCTTGGATG 1332

Db 300 cacactttgtgtacatagttacccttttaactcagtggttatctaatagcttcaactca 359  
QY 1333 CACACTTTGTGTGTACATAGTACCTTTAACTCAGTGGTTATCTTAATAGCTCTTAACATCA 1392

Db 360 ttaaaaaactcccaagccttcccaaaaaacagtgccccaccctgtatacatattttataaa 419  
QY 1393 TTAATAAAACTCCAAAGCCTTCCCAAAAACAGTGCACCCTGTATACATTTTATTAAA 1452

Db 420 aaaaatgaatcttattggtataaactggtatgataatgcttcttatgaatgatgttg 479  
QY 1453 AAAATGTAATGCTTATG-TATAAACATG-TATGTAATATAGCTTTCTATGAATGATGTTG 1510

Db 480 atttaaanatacatatataaaatgatggng 514  
QY 1511 ATTTAAA-TATAATACATATTAATAATGATGGAG 1544

# RESULT

2

LOCUS

202008.r1

587439 5'

AA132640

AA132640

Q1694147

EST

Human.

ORGANISM

Homo sapiens

Eukaryotae; mitochondria

Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 390)

Hillier L., Clark N., Dubucque T., Elliston K., Hawkins M.,

Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,

Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,

Waterston R., Williamson A., Wohlmann P. and Wilson R.

WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilson RK

WashU-Merck EST Project

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham.

FEATURES

Source

1..390

/organism="Homo sapiens"

/note="Organ: colon; Vector: pBluescript SK-; Site.1:

EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. T-84 colonic epithelial cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'

CTCGAGCTTTTCTTTTCTTTT 3'.

/clone="587439"

/clone.lib="Stratagene colon (#937204)"

/lab\_host="SOLR cells (kanamycin resistant)"

<1..>390

BASE COUNT 111 a 81 c 82 g 115 t 1 others

# ORIGIN

Query Match 24.3%; Score 380; DB 9; Length 390;  
Best Local Similarity 99.2%; Pred. No. 0.00e+00;  
Matches 387; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 gaagcgttgagtttttaagtgttagagcttagatattagataaacactgagttaaagta 60  
Cp 1413 GAAGCCTGGAGTTTTTTTAAAGTGTATAGAGCTATTAGATAAACCACTGAGTTAAAGTA 1354

Db 61 actatgtacacacaaagtgcaatccaaagagcagatagcagcagcagaagctcttaag 120  
Cp 1353 ACTATGTACACACAAAGTGTGCA-TCCAGAGGCGATATCAGCAGCAGAGTCTTTAAGG 1295

Db 121 cttgtacaccaggaagaagatgcattccttgccttgccttgccttgccttgccttgccttgc 180  
Cp 1294 CTTGTACACCAGGAAGAAGATGCATCCTCTTGCCTTGTGGCAATCATTTTCTCTTGA 1235

Db 181 aacagggcagcttcacctggcaccctgctccttcaagctggtgattgctcggtatg 240  
Cp 1234 AACAGGCGAGCTTCACCTGGGACCCCTGGCTTCAAGGCTGCTGATTCGCGATAG 1175

Db 241 tgattccacgttgggtgttgcattgcagagttgattgagagtgctcctctcttctcttct 300  
Cp 1174 TGAATCCAGTTGTTGTTTCATGCAGAGTTGTATGAGAGTCTCTCTCTTTCTTCT 1115

Db 301 ttaaaagaagttcttcttcttgaagaaatccgatacatatcacgcaactaatattgcaacc 360  
Cp 1114 TTAAGAAGTCTTCTTTTGAAGAAATCCGATACATATACAGCAACTAATATTGCAACC 1055

```

Db 361 agagctccctgaatgagtcagtcacacaca 390
|||||
Cp 1054 AGAGCTCCCTGAATGAGTCCAGTCAACACA 1025
|||||

RESULT 3
ID HS605336 standard; RNA; EST; 415 BP.
AC W24605;
NI 91301591
DT 09-MAY-1996 (Rel. 47, Created)
DT 07-MAR-1997 (Rel. 51, Last updated, Version 2)
DE ZB63f08.r1 Soares fetal lung NBHL19W Homo sapiens cDNA clone 308295
DE 5'
DE 3'
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP 1-415
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Tan F., Trevasakis E.,
RA Waterston R., Williamson A., Wohlmann P., Wilson R.;
RA "WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estevenson.wustl.edu This clone is available royalty-free through
CC LLNL; contact the IMAGE Consortium (infoimage.llnl.gov) for
CC further information. Insert Length: 773 Std Error: 0.00 Seq primer:
CC mob.REGA+ET.
FH Key
FH Location/Qualifiers
FH 1..415
FH source
FH 1..415
FH /organism="Homo sapiens"
FH /note="Organ: lung; Vector: pT73D (Pharmacia) with a
FH modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
FH strand cDNA was primed with a Not I - oligo(dT) primer
FH [5'-TGTTACCAATCTGAAGTGGGCGGCGCAATTTTTTTTTTTT-3'],
FH double-stranded cDNA was size selected, ligated to Eco RI
FH adapters (Pharmacia), digested with Not I and cloned into
FH the Not I and Eco RI sites of a modified pT73 vector
FH (Pharmacia). Library went through one round of
FH normalization to a Cot = 5. Library constructed by Bento
FH Soares and M.Patima Bonaldo. This library was constructed
FH from the same fetus as the fetal heart library, Soares
FH fetal heart NBHL19W."
FH /clone="308295"
FH /clone.lib="Soares fetal lung NBHL19W"
FH /dev.stage="19 weeks"
FH /lab_host="DH10B (ampicillin resistant)"
FH <1..415
FH mrna
SQ Sequence 415 BP; 124 A; 87 C; 86 G; 116 T; 2 other;

Query Match 24.0%; Score 376; DB 87; Length 415;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 408; Conservative 0; Mismatches 3; Indels 5; Gaps 4;

Db 1 gagttctgattataaacaccactggagcgtgtgtgactgagctcattcaggggctc 60
|||||
Qy 994 GAGTTTCTGATTATAAACACCCTGGAGCGATGTGTGACTGGACTCATTCAGGGAGCTC 1053
|||||
Db 61 tgggtgcaataggctgtgtatgtatgatttcttcaagaagaactctcttta 120
|||||
Qy 1054 TGCTTCAATATTAGTTGCTGTATGATGATCGATTCTTCAAGAAGAAGACTCTTTA 1113
|||||
Db 121 aagaagaagaaggaggactctctatacaactctgcatgaacacccaaactcggaaac 180
|||||
Qy 1114 AAGAAAAGAAAGAGGAGGAGTCTCTATCAACTCTGCATGAAACACCAACACTGGGAATC 1173
|||||
Db 181 actatcnccagcgcaatcaccagccttgaaaggcagcaggggtgcccaggctgagcctggcc 240
|||||

Qy 1174 ACTATC-C-GAGCAATCACCAGCTTGAAGGCGAGGCTGCCAGGTGAGCTGGCC 1230
|||||
Db 241 tgtttctaaaggaataatgctgcacaaag-aagaggatgcatcttcttctctggtgta 299
|||||
Qy 1231 TGTTTTCTAAAGGAAATGATTGCCAAGCAAGAGGATGCAATCTTCTTCTCTGTTGTA 1290
|||||
Db 300 caagccttttaaagacttctgctgtctaatgctctcttggatgcacacttgggtgtaca 359
|||||
Qy 1291 CAAGCTTTAAAGACTTCTGCTGCTGATA-TGCCCTTGGATGCACACTTTGTGTGTACA 1349
|||||
Db 360 tagttaaccttaactcagtggttatctataatagctctaaactcattaaaaaactcc 415
|||||
Qy 1350 TAGTACCTTTAACTCAGTGTGTATCTAATAGCTCTAAACTCATTAATAAAACTCC 1405
|||||

RESULT 4
LOCUS AAL32534 376 bp mRNA EST 27-NOV-1996
DEFINITION Z020d08.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
587439.3'.
ACCESSION AAL32534
NID 91694041
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 285.
Location/Qualifiers
1..376
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site.1:
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3'"
/clone.lib="587439"
/clone.lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
complement(<1..>376)
BASE COUNT 107 a 78 c 83 g 108 t
ORIGIN
Query Match 22.9%; Score 359; DB 9; Length 376;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 373; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Db 1 gagcgtgtgttgaactgagctcattcaggagctcctggttgcgaatattagttcgtata 60
|||||
Qy 1019 GAGCGATGTGTGACTGGACTCATTCAGGAGCTCTGGTTCGAATATTAGTCTGTATA 1078
|||||
Db 61 tgtatcggtattcttcaagaagaactctctttaaagaagaagaagaggagactctca 120
|||||
Qy 1079 TGATCGGATTTCTTCAAGAAGAAGAACTCTTTTAAAGAAGAAGAAGAGGAGGACTCTCA 1138
|||||

```





## JOURNAL COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:313587

Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 375.

## FEATURES

source  
 Location/Qualifiers

1..385

/organism="Mus musculus"

/strain="C57/B16"

/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Cloned unidirectionally. Primer: Oligo dt. Average

insert size: 1.0 Kb; Uni-ZAP XR vector; -5' adaptor

sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3'

/clone="519739"

/clone\_lib="Stratagene mouse kidney (#937315)"

/sex="females"

/dev\_stage="4 weeks"

/lab\_host="SOLR (kanamycin resistant)"

<1..>385

BASE COUNT 95 a 84 c 89 g 117 t

ORIGIN

Query Match 16.5%; Score 259; DB 33; Length 385;

Best Local Similarity 84.2%; Pred.No. 0.00e+00;

Matches 325; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

Db 1 atgagttatggagaatctctgtctgttactttaattgtctgtcattcgaattctttgtc 60

QY 555 ATTATCTTGAGAAACCCCTGCTGTTACTGTGTAACCTTTTACACAAATCTTTATC 614

Db 61 ggcacccctacacagccattcaaacgcgtcgagccctttttgttcggagtctca 120

QY 615 AGTAATACTACATAGCCACTATTACAAAGCCATTGGAACTTTTATTTGGTGCAGCT 674

Db 121 gctagtcagtccttcgctgcagtcgtagtactatagtcaggttgcggccg-acttc 179

QY 675 GCTAGTCAGTCCCTGACTGACATGCAAGTATTCATAGGCAGACTCGGCCCTCACTTC 734

Db 180 ttggctactgttaaccagactggtcacaatacaactcagtcagtgatggctatattgaggac 239

QY 735 TTGGATGTTTGTATCCAGATTGGTCAAAATCACTCGAGGATGGTTACATTGAATAC 794

Db 240 tacatatgcaagggaatgaagagaagtcgaaggagggcaggttgcctttcttactcggga 299

QY 795 TACATATGTCGAGGGAATGAGAAAGAGTTAAGGAAGCAGGTTGCTCTTATTCAGGC 854

Db 300 cactctcattctctatgtactgcagtcgtttgttcgcacttattctcgaagcagagtg 359

QY 855 CACTCTTCGTTTCCATGTACTGCATGCTGTTGTGGCATTTATCTTCAAGCCAGGATG 914

Db 360 aaggagagactgggcaagactcttcag 385

QY 915 AAGGGAGACTGGGCAAGACTCTTACG 940

RESULT 9 AA210789 542 bp mRNA EST 31-JAN-1997

LOCUS zr-90a08.r1 Soares NbrHTGBC Homo sapiens cDNA clone 682982 5'

DEFINITION AA210789

ACCESSION 91809443

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

## REFERENCE

AUTHORS

## TITLE

JOURNAL

COMMENT

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo, 1 (bases 1 to 542)  
 Hillier, L., Clark, N., Dubouche, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, P., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 414.

## FEATURES

source

Location/Qualifiers

1..542

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand

was prepared from germinal B-cells (flow-sorted from

tonsils) provided by Dr. Louis Staudt of the NCI, and was

then primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTCTTTTCTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

/clone="682982"

/clone\_lib="Soares NbrHTGBC"

/lab\_host="DH10B"

<1..>542

BASE COUNT 162 a 97 c 117 g 164 t 2 others

ORIGIN

Query Match 14.6%; Score 229; DB 22; Length 542;

Best Local Similarity 95.4%; Pred.No. 4.47e-298;

Matches 308; Conservative 0; Mismatches 2; Indels 13; Gaps 13;

Db 220 ttgtgtctcccatcatatttaatatgtattattattataataaatacaatcattcagaaa 279

Cp 1551 TTGTGTTCTCCCATCATTTTATATGATATATATTTAAATCAACATCATCATAGAAA 1492

Db 280 gcattattacacatgtttttatatacaataagcattacatttttttaataaaatgtatcacagg 339

Cp 1491 GCATATTACATACATGTTTATACATAAAGCATTTACATTTTAAATAAATAATGTATACAGG 1432

Db 340 tggggcactgttttctgtgaagccttgaggttttttttaataaggttttagagcttagat 399

Cp 1431 TGGGGCACTGTTTGGTGGGAAGGCTTGGAGTTTTTTT-AAATGAGTTAGAGCTATTAGAT 1373

Db 400 aaccactgagttaaagggttaactatggttacacacaaaaggtgtgccatcccaagaagcaca 459

Cp 1372 AACCACTGAGTTAAAGG-TAACTATG-TACACACAAAG-TGTGC-ATCCAAGA-GGCATA 1318

Db 460 gcagcagcagaagctctttaaaggcttgaccnccagggaggaagaggtatcatccctcctt 519

Cp 1317 TCAGCAGCAGAA-GTCTTTAAAGGCTTGTAC-ACCAGG-AAGAAG-ATGCATCTTC-TT 1263

Db 520 ggccttggtggcaatcattttcc 542

Cp 1262 G-CCTTG-TGGCAATCATTTTCC 1242

RESULT 10





```

RESULT 12
LOCUS   human STS WI-15758.      STS          31-MAY-1996
DEFINITION
ACCESSION   G3282
NID         G3282
KEYWORDS   STS sequence; primer; sequence tagged site.
SOURCE     human STSs derived from sequences in dbEST and the Unigene
           collection.
ORGANISM   Homo sapiens
           Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
           Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 501)
AUTHORS   Hudson,T.
TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
JOURNAL   Mapped STSs
COMMENT   Unpublished (1995)
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TATTCAGTATTGGCAAAATGAAGC
Primer B: TACTTTGAGGTGTACAGATTAAAGCA
STS size: 129
PCR profile:
  Denaturation:
    Annealing: 56 degrees C
  Polymerization:
    PCR Cycles: 35
  Thermal Cycler:
Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: each 4 nM
  Taq Polymerase: 0.025 units/ul
  Total Vol: 20 ul
Buffer:
  MgCl2: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 9.3
Derived from dbEST (genbank accession R45479).
FEATURES             Location/Qualifiers
     source           1..501
                     /organism="Homo sapiens"
                     /map="188.1 cR from top of Chr5 linkage group"
     STS              23..151
     primer_bind      23..46
     primer_bind      133 a      88 c      117 g      157 t      6 others
BASE COUNT          133 a      88 c      117 g      157 t
ORIGIN
Query Match          7.6%; Score 119; DB 43; Length 501;
Best Local Similarity 84.9%; Pred. No. 6.68e-129;
Matches 202; Conservative 0; Mismatches 26; Indels 10; Gaps 10;
Db 255 gctgtatgtatgctgggtttcttcaagaaggagactctttcaaggaggaagaaagg 314
Qy 1071 GCTGTATGTATGCGGATTCTTCAGAAGAAAGACTCTTTTAAG-AAAG-AAAGAGG 1128
Db 315 gggggctctatacaactctggcatggagaaacccnaccnaccactggggttcattccagg 374
Qy 1129 AGGA-CTCTCATCAACTCTG-CATG-AAACACCAACAACTGGGAAT-CACATCCGAG- 1183

Db 375 caatacacagccttgaaaggcgggcgccgggggtgagggcctggcctgtttttaa 434
Qy 1184 CAATCACAGCCTTGAAGG-CAGCAGGCTGCCAGG-TGAAGTGG-CCTGTTTCTAA 1240
Db 435 ggggaaatgggtgccacagncaggggggtgctcttcttccctccgggggtacagcctt 492
Qy 1241 AGGAATATGATTCACCAAGGCAAGAGGATCATCTTCTTCTCTGGGTGTACAGCCTT 1298

RESULT 13
ID       MM115742 standard; RNA; EST; 313 BP.
AC       AA237956;
NI       G1862038
DT       06-MAR-1997 (Rel. 51, Created)
DE       mx78g01.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'
DE       similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN
DE       .
DE       .
KW       EST.
OS       Mus musculus (house mouse)
OC       Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC       Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC       Mus.
RN       [1]
RP       1-313
RA       Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,
RA       Dubuque T., Geisels S., Kucaba T., Lacy M., Le M., Martin J.,
RA       Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,
RA       Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,
RA       Waterston R.;
RT       "The WashU-HMI Mouse EST Project";
RL       Unpublished
CC       Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project
CC       Washington University School of Medicine 444 Forest Park Parkway,
CC       Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810
CC       Email: mouseest@watson.wustl.edu This clone is available
CC       royalty-free through LLNL; contact the IMAGE Consortium
CC       (info@image.llnl.gov) for further information. MGI:426056 Possible
CC       reversed clone: similarity on wrong strand Seq primer: -28m13 rev2
CC       ET from Amersham High quality sequence stop: 57.
FH       Key
FH       Location/Qualifiers
FT       source
FT       1..313
FT       /organism="Mus musculus"
FT       /note="Vector: pT73D-Pac (Pharmacia) with a modified
FT       polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
FT       was primed with a Not I - oligo(dT) primer [5',
FT       TGTACCAATCTGAGTGGGAGCGCGCGCAATCTTTTCTTTT 3']];
FT       TGTACCAATCTGAGTGGGAGCGCGCGCAATCTTTTCTTTT 3']];
FT       double-stranded cDNA was ligated to Eco RI adaptors
FT       (Pharmacia), digested with Not I and cloned into the Not I
FT       and Eco RI sites of the modified pT73 vector. Library
FT       constructed and normalized by Bento Soares and M.Fatima
FT       Bernaldo.
FT       /clone="692496"
FT       /clone_lib="Soares mouse NML"
FT       /tissue_type="Liver"
FT       /lab_host="DH10B"
FT       mRNA
FT       <1..>313
SQ       Sequence 313 BP; 83 A; 58 C; 85 G; 87 T; 0 other;

Query Match          5.6%; Score 88; DB 91; Length 313;
Best Local Similarity 66.5%; Pred. No. 1.86e-83;
Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;
Db 3 gagtccactgtacactcactcactgtgtgtgttagtcacacactcagaaagcc-ac 61
Cp 1040 GAGTCCAGTCAACACATCGCTCCAGTGGTGTATATACAGAACTCGAAGAGCCAC 981
Db 62 atatatgaaagaaagcaatgagccaaactggagcaatggtcgtgaagcctggccaatt 121
Cp 980 ATAAATGGATACGCGCAACAAGACCAATTCAGATGTGGGGCGTAAGA-GTCTTCCAGT 922
Db 122 ttctttaactctgggtcttgagggttaagttgcacacacaggattgagttacttggggattg 181

```

```

Cp 921 CTCCTTCATCTGCTTGAAGATGAAGTGCACACAGCATGCGAGTACATGGAAACG 862
D 182 agaattccgggttggaagcccaactggcctcttcttgattccct-caatcccttggc 240
Cp 861 AAGAGTGGCGTGAATAGAGGACACACCTGCTTCCCTTAATCTTTCTGATTCCTCGAC 802
D 241 ataaagagttctcatatggcataatgaagtgagttgattttgaacagttctcggttataga 300
Cp 801 ATATGTAGTATCAATGTAACTCACTGCTGAGTGTATTTTGACCAATCTGATCACA 742
D 301 gatgccagaatg 313
Cp 741 CATCCAAGAGTG 729

RESULT 14
LOCUS AA237956 313 bp mRNA EST 03-WAR-1997
DEFINITION mx78901.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'
similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN
;
ACCESSION AA237956
NID G1862038
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Eukarya; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 313)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:426056
Possible reversed clone; similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 57.
Location/Qualifiers
1..313
/organism="Mus musculus"
/vector="p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTGAGTGGAGCGCGGCGGAGTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
/clone="692496"
/clone_lib="Soares mouse NML"
/tissue_type="liver"
/lab_host="DH10B"
<1..>313
83 a 58 c 85 g 87 t

BASE COUNT
ORIGIN
Query Match 5.6%; Score 88; DB 66; Length 313;
Best Local Similarity 66.5%; Pred. No. 1.86e-83;

FEATURES
source
1..564
/organism="Homo sapiens"
/clone="p7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTGAGTGGAGCGCGGCGGAGTCTTTTCTTTT 3'];
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by M. Fatima
Bonaldo. This library was constructed from the same fetus
as the fetal lung library, Soares fetal lung NBHL19W."
/clone="327113"
/clone_lib="Soares fetal heart NBHL19W"
/sex="unknown"

```

```

Matches 208; Conservative 0; Mismatches 102; Indels 3; Gaps 3;
D 3 gagtccacttgacgtcactccagtggtttgttagtcagacactcagagaagcc-ac 61
Cp 1040 GAGTCCAGTCAACACATCGCTCCAGTGTGTTTATATCAGAAACTCGAGAAAGCCAC 981
D 62 atatatgaaaaaacaatgagcccaactgagcaaatggtcgttaagaagccctggccaatt 121
Cp 980 ATAAATGATACGCGACACAGACCAATTCAGTGTGGGCGGTAGA-GTCTTGCCAGT 922
D 122 ttcttttaattggcttgggttaagttaattgacacagagattgagttacttggggattg 181
Cp 921 CTCCTTCATCTGCTTGAAGATGAAGTGCACACAGCATGCGAGTACATGGAAACG 862
D 182 agaatttccgggttggaagcccaactggcctcttcttgattccct-caatcccttggc 240
Cp 861 AAGAGTGGCGTGAATAGAGGACACACCTGCTTCCCTTAATCTTTCTGATTCCTCGAC 802
D 241 ataaagagttctcatatggcataatgaagtgagttgattttgaacagttctcggttataga 300
Cp 801 ATATGTAGTATCAATGTAACTCACTGCTGAGTGTATTTTGACCAATCTGATCACA 742
D 301 gatgccagaatg 313
Cp 741 CATCCAAGAGTG 729

RESULT 15
ID HS942332 standard; RNA; EST; 564 BP.
AC W30942;
NI G1311934
DT 13-MAY-1996 (Rel. 47, Created)
DT 07-WAR-1997 (Rel. 51, Last updated, Version 2)
DE zc64f09.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
DE 327113 5' similar to WP:T28D9.3 CE02068 ;
KW EST.
OS Homo sapiens (human)
OC Eukaryota; Eukarya; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP 1-564
RA Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
RA Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
RA Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
RA Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., Wilson, R.;
RT "The WashU-Merck EST Project";
RL Unpublished.
CC Contact: Willson RK WashU-Merck EST Project Washington University
School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estewatson.wustl.edu This clone is available royalty-free through
CC LLM; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:
CC mob.REGA+ET High quality sequence stop: 336.
Key Location/Qualifiers
1..564
/organism="Homo sapiens"
/clone="p7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTGAGTGGAGCGCGGCGGAGTCTTTTCTTTT 3'];
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by M. Fatima
Bonaldo. This library was constructed from the same fetus
as the fetal lung library, Soares fetal lung NBHL19W."
/clone="327113"
/clone_lib="Soares fetal heart NBHL19W"
/sex="unknown"

```

```
FT /dev_stage="19 weeks"
FT /lab_host="DH10B (ampicillin resistant)"
FT <1>>564
SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;

Query Match      4.9%; Score 76; DB 88; Length 564;
Best Local Similarity 67.8%; Pred. No. 1.97e-66;
Matches 158; Conservative 0; Mismatches 73; Indels 2; Gaps 2;

Db 251 agccagtcctttcacagacattgcacaagtgccatagggcgccctgcctcactctttg 310
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 678 AGTCAGTCCCTGACTGACATGGCCAGTATTCAATAGGCAGACTCGGCGCTCACTCTTTG 737
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 311 aagtgtctgcaacctgtatttcagccagatcaactgctctgaaggctacattcagaacta 370
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 738 GA-TGTTTGTGATCCAGATTGGTCAAAAATCAACTGCAGCGATGCTTACATTGAATACTA 796
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 371 cagatgcagagtgatgacagcaaatccaggangccaggaaagtctctctcttgcca 430
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 797 CATATGTCGAGGGAATCAGAAAGATTAGGAAGCGAGGTGTCTCTTATTTCAGGCCA 856
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 431 tgcctctcttccatgtacactatgtggaatttggtgctatancctgcaggcc 483
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 857 CTCCTCGTTTTCATGTACTGCATGCTGTT-TGTGGCACTTTATCTTCAAGCC 908
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: Thu Nov 6 11:35:37 1997  
Job time : 499 secs.

21	90	4.2	470	7	R34478	Encoded by Hepatitis	4.11e+01
22	90	4.2	496	21	W20801	H. pylori inner membr	5.11e+01
23	88	4.2	628	5	R2575	ABF-A from A. niger.	5.62e+01
24	88	4.2	758	6	R28954	Vitamin K dependent c	5.62e+01
25	88	4.2	1063	13	R73916	Rubella virus strain	5.62e+01
26	88	4.2	1141	6	R31961	Human cardiac CGI PDE	5.62e+01
27	90	4.2	1189	11	R56496	TATA-binding protein-	4.11e+01
28	88	4.2	1353	19	R92925	Murine adenylate cycl	5.62e+01
29	90	4.2	1463	12	R63575	Group 1 phospholipase	4.11e+01
30	90	4.2	1463	16	R85143	Bovine phospholipase	4.11e+01
31	86	4.1	270	18	R93586	Low density lipoprote	7.66e+01
32	86	4.1	273	18	R93587	Low density lipoprote	7.66e+01
33	87	4.1	289	12	R66796	Novel mouse proteogly	6.56e+01
34	87	4.1	310	17	R87000	Human syndecan-1.	6.56e+01
35	87	4.1	311	12	R66793	Novel mouse proteogly	6.56e+01
36	87	4.1	311	10	R55276	Syndecan protein.	6.56e+01
37	87	4.1	311	17	R87001	Mouse syndecan-1.	6.56e+01
38	87	4.1	389	3	R15428	3-acylation enzyme.	6.56e+01
39	87	4.1	428	7	R39264	Murine somatostatin r	6.56e+01
40	87	4.1	643	18	R37735	Bacillus thuringiensis	6.56e+01
41	87	4.1	643	1	R31462	67-kD protein toxin.	6.56e+01
42	87	4.1	1063	15	R79049	Infectious rubella vi	6.56e+01
43	87	4.1	1353	19	R93938	Human restrictin.	6.56e+01
44	86	4.1	1876	14	R72853	S. cerevisiae FKSI.	7.66e+01
45	86	4.0	262	8	R42502	A. rhizogenes NIAES17	8.93e+01

## ALIGNMENTS

	RESULT	1
ID	W20304	standard; Protein; 193 AA.
AC	W20304;	
DC	10-JUN-1997	(first entry)
DE	H. Pylori cytoplasmic protein 24407533.tat.	
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification;	
DD	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;	
CC	duodenal ulcer disease; chronic gastritis; diagnosis; envelope.	

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.	
1	109	5.1	129	21	W20304	H. pylori cytoplasmic	1.86e+00
2	109	5.1	223	21	W20620	H. pylori cytoplasmic	1.86e+00
3	105	5.0	682	19	W04359	S. pneumoniae penicil	3.63e+00
4	96	4.5	397	18	W04217	Rat dorsal root gangl	1.58e+01
5	93	4.4	317	19	W02661	G-protein coupled hum	2.56e+01
6	93	4.4	317	16	R38699	G-protein coupled hum	2.56e+01
7	94	4.4	418	7	R39283	Human somatostatin re	2.18e+01
8	94	4.4	563	10	R32463	Pea proton transport	2.18e+01
9	94	4.4	1073	10	R35272	Alpha subunit of inte	2.18e+01
10	91	4.4	273	5	R38821	Alpha 6A integrin sub	2.18e+01
11	92	4.3	276	21	W20612	H. pylori protein.	3.00e+01
12	91	4.3	1091	5	R38832	Alpha 6B integrin sub	3.51e+01
13	92	4.3	1539	21	W13905	Human SWC protein.	3.00e+01
14	91	4.3	3011	7	R34468	Encoded by full-length	3.51e+01
15	88	4.2	170	21	W20704	H. pylori transmembra	5.62e+01
16	90	4.2	205	21	W20288	H. pylori transporter	4.11e+01
17	89	4.2	250	6	R28365	Human LFA-3 protein.	4.81e+01
18	90	4.2	270	21	W20935	H. pylori transporter	4.11e+01
19	90	4.2	470	7	R34467	Encoded by Hepatitis	4.11e+01
20	90	4.2	470	7	R34476	Encoded by Hepatitis	4.11e+01

```

Db 14 ifdmshlkyefkalfghndskqkfgyhyqsgisrnekiqfyneilktpaq 73
QY 1 MFDRTRLPYVALDVLCLASMPAVLKLGQIYFQFGFCCKDINSINTPYDSTAATVL 60
Db 74 eisdalalefagai-eqklfdgrhlhnmvmfakdkhyqnyfihiasaalhselqvlcefl 132
QY 61 ILV-GVGLPVSSIIIGETLSVYCNLL-HSNSFISNNYIATIKYKGTFLFGAASQS-LT 117
Db 133 gttky 137
QY 118 DIARY 122

RESULT 2
ID W20620 standard; Protein; 229 AA.
AC W20620.
DE 14-JUL-1997 (first entry)
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR N-PSDB: T67873.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PS infection, and to detect Helicobacter
PS Claim 61; Page 1047; 1481pp; English.
CC This sequence represents a H. pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 229 AA;

Query Match 5.1%; Score 109; DB 21; Length 229;
Best Local Similarity 20.8%; Pred. No. 1.86e+00;
Matches 26; Conservative 36; Mismatches 59; Indels 4; Gaps 4;

Db 21 ifdmshlkyefkalfghndskqkfgyhyqsgisrnekiqfyneilktpaq 80
QY 1 MFDRTRLPYVALDVLCLASMPAVLKLGQIYFQFGFCCKDINSINTPYDSTAATVL 60
Db 81 eisdalalefagai-eqklfdgrhlhnmvmfakdkhyqnyfihiasaalhselqvlcefl 139
QY 61 ILV-GVGLPVSSIIIGETLSVYCNLL-HSNSFISNNYIATIKYKGTFLFGAASQS-LT 117
Db 140 gttky 144
QY 118 DIARY 122

RESULT 3
ID W04359 standard; Protein; 682 AA.
AC W04359;

```

```

DT 03-DEC-1996 (first entry)
DE S. pneumoniae penicillin binding protein 1A soluble variant.
KW penicillin binding protein; PBP 1A; bifunctional protein;
KW transglycosylase; transpeptidase; identification; assay; inhibitor;
KW antibiotic resistant; bacteria; soluble variant; protein structure;
KW X-ray crystallography; determination.
OS Streptococcus pneumoniae.
PN GB2290792-A.
PD 10-JAN-1996.
PF 29-JUN-1995; 013306.
PR 01-JUL-1994; IN-000580.
PR 24-NOV-1994; SE-004072.
PA (ASTR ) ASTRA AB.
PI Balganesht TS, Town CM;
DR WPI: 96-042232/05.
DR N-PSDB: T08027.
PT Sol. derivs. of bifunctional penicillin binding protein (BPBP) -
PT opt. lack transglycosylase activity, useful to identify and assay
PT for antibodies or cpds. which bind BPBPs
PS Claim 2; Pages 65-68; 108pp; English.
CC The present sequence is a soluble variant of the S. pneumoniae
CC penicillin binding protein (PBP) 1A, which lacks the 1st 38 amino
CC acid residues of the wild type protein. Wild type PBP is a
CC bifunctional protein, which binds the cell membrane when expressed
CC in a bacterial cell, having transglycosylase and transpeptidase
CC activities. The variant protein (NCIMB 40665) in conjunction with
CC a labelled anti-bifunctional PBP monoclonal antibody, can be used
CC to identify and assay for cpds. which bind bifunctional PBP. Such
CC cpds., as inhibitors of bifunctional PBP have a potential use in
CC therapeutic cpds. which inhibit the growth of antibiotic resistant
CC bacteria. The soluble variant may also be used in X-ray
CC crystallography.
SQ Sequence 682 AA;

Query Match 5.0%; Score 105; DB 19; Length 682;
Best Local Similarity 28.6%; Pred. No. 3.63e+00;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 371 vynywdrgyfgnltqyvalqgarnvavetlnkvlnraktf1-ngligidypsishyana1 428
QY 32 IYFQRFCKDINSINYPHDSTAATVLLVGLGVSSIIIGETLSV-FCNLLHSNFS 90
Db 429 ssn 431
QY 91 ISN 93

RESULT 4
ID W04217 standard; Protein; 397 AA.
AC W04217;
DT 23-NOV-1996 (first entry)
DE Rat dorsal root ganglion P2x receptor.
KW ATP P2x receptor; purinoceptor; ligand-gated ion channel; agonist;
KW antagonist; epilepsy; cognition; emesis; pain; asthma;
KW peripheral vascular disease; hypertension; irritable bowel syndrome;
KW premature ejaculation; cystitis; therapy.
OS Rattus sp.
PN W09533048-A2.
PD 07-DEC-1995.
PF 24-MAY-1995; E01968.
PR 27-MAY-1994; GB-010664.
PR 09-FEB-1995; GB-002480.
PA (GLAX ) GLAXO GROUP LTD.
PI Buell GN, Valera S;
DR WPI: 96-030561/03.
DR N-PSDB: T33854.
PT DNA encoding ATP P2x receptors of the purinoceptor family - for
PT screening cpds. useful in treating epilepsy, cognition, emesis,
PT pain, asthma, peripheral vascular disease, hypertension, etc.
PS Claim 1; Fig 3; 82pp; English.
CC The amino acid sequence (W04217) of rat dorsal root ganglion P2x
CC receptor was deduced from a cDNA clone (T33854) isolated from a
CC cDNA library. P2x receptors (see also W04215-16 and W04219) are

```



Db 94 mlg 96  
QY 107 LFG 109

RESULT 8

ID R53463 standard; Protein; 563 AA.  
AC R53463;  
DT 18-JAN-1995 (first entry)  
DE Pea proton transport adenosine triphosphatase.  
KW Proton transport adenosine triphosphatase; H+ATPase; pea.  
OS Pisum sativum.  
PN J06113862-A.  
PD 26-APR-1994.  
PF 29-SEP-1992; 260333.  
PR 29-SEP-1992; JP-260333.  
PA (YAWH) NIPPON STEEL CHEM CO.  
DR (YAWA) NIPPON STEEL CORP.  
WI: 94-172755/21.  
N-PSDB; Q63601.

PT Proton transport adenosine triphosphatase protein and its coding  
PT gene - for preparation of pure proton transport adenosine  
PT tri-phosphatase.  
PS Claim 2; Page 6-7; 8pp; Japanese.  
CC This sequence is a proton transport adenosine triphosphatase  
CC protein from the plasma membrane of Pisum sativum.  
SQ Sequence 563 AA;

Query Match 4.4%; Score 94; DB 10; Length 563;  
Best Local Similarity 27.0%; Pred. No. 2.18e+01;  
Matches 20; Conservative 20; Mismatches 31; Indels 3; Gaps 3;

Db 418 fpkvfgvatleknaahdfklsaylvqstisqalifvtrsgwsvyverglllvaafi 477  
QY 178 YCMLFVALVLAQRKMGDWARLRLPT-LQFLGVAVS-IYVGLSRVSVDLVLTGILI 234

Db 478 vagliatlaiavvas 491  
QY 235 QGALVALVAVVVS 248

RESULT 9

ID R55272 standard; Protein; 1073 AA.  
AC R55272;  
DT 27-JAN-1995 (first entry)  
DE Alpha subunit of integrin cell surface receptor.  
KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;  
KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;  
KW behaviour; signal transduction; receptor.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT misc\_binding 230..238  
FT /note- "Sequence weakly homologous with cation  
FT binding site of other integrin alpha  
FT subunits."  
FT misc\_binding 324..332  
FT /note- "Putative cation binding site."  
FT misc\_binding 385..394  
FT /note- "Putative cation binding site."  
FT misc\_binding 441..449  
FT /note- "Putative cation binding site."  
FT misc\_feature 899..903  
FT /note- "This sequence closely resembles the  
FT cleavage site of other integrin O chains."  
FT misc\_feature 1040..1044  
FT /\*tag= f  
FT /note- "This sequence is conserved in all but the  
FT Drosophila PS2 alpha subunit"  
FN Drosophila PS2 alpha subunit  
PD 14-JUN-1994.  
PF 19-FEB-1987; 016552.  
PR 19-FEB-1987; US-016552.  
PR 04-JAN-1989; US-293384.

01-OCT-1990; US-591105.  
PA (KAJI/) KAJIJI S.  
PA (QUAR/) QUARANTA V.  
PI Kajiji S., Quaranta V.  
DR WPI: 94-191533/23.  
DR N-PSDB; Q65673.

PT Diagnosing presence of abnormal epithelial tissue in vitro -  
PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein  
PS Example 5; Figure 6; 34pp; English.  
CC Integrins are heterodimers comprised of alpha and beta subunits, that  
CC are non-covalently associated transmembrane glycoproteins. 11 alpha  
CC chains and 6 beta chains have been recognised in man. Each alpha  
CC subunit tends to associate with only one type of beta subunit but  
CC there are several exceptions. Integrins mediate (in part) the  
CC interaction of cells with the extracellular matrix, forming a link  
CC between the extracellular matrix and the cytoskeleton. They may  
CC transmit signals from the extracellular to the intracellular  
CC environment, affecting cell behaviour. This sequence is the alpha6  
CC subunit of an alpha6 beta4 integrin.  
SQ Sequence 1073 AA;

Query Match 4.4%; Score 94; DB 10; Length 1073;  
Best Local Similarity 25.3%; Pred. No. 2.18e+01;  
Matches 20; Conservative 23; Mismatches 29; Indels 7; Gaps 7;

Db 997 vrvtvfpkt-vagysgvpmilvailagilmallvflmkcgffk-rnk-kdhydat 1053  
QY 208 VAVSIYVGLSRVSDYK-HHWSDVLTLGIQCALV-AILVAV-VVSDPKERTSFKERKEED 264

Db 1054 yhkaeihagpadkerltsd 1072  
QY 265 SHYT-LHETPTTGNHYPSN 282

RESULT 10

ID R28821 standard; Protein; 1073 AA.  
AC R28821;  
DT 23-MAR-1993 (first entry)  
DE Alpha 6A integrin subunit.  
KW Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;  
KW extracellular matrix; cytoskeleton; heterodimer; laminin receptor;  
KW immunoprecipitation; JAR; chorioncarcinoma.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Domain 1012..1037  
FT /label- Transmembrane\_domain  
FT Peptide 1..23  
FT /label- Signal\_peptide  
FT Protein 24..1073  
FT /label- Mature\_protein  
FT Modified\_site 223  
FT /label- Glycosylation\_site  
FT Modified\_site 284  
FT /label- Glycosylation\_site  
FT Modified\_site 370  
FT /label- Glycosylation\_site  
FT Modified\_site 513  
FT /label- Glycosylation\_site  
FT Modified\_site 731  
FT /label- Glycosylation\_site  
FT Modified\_site 748  
FT /label- Glycosylation\_site  
FT Modified\_site 891  
FT /label- Glycosylation\_site  
FT Modified\_site 927  
FT /label- Glycosylation\_site  
FT Modified\_site 958  
FT /label- Glycosylation\_site  
FT Binding\_site 230..238  
FT /note- "Putative cation binding domain"  
FT Binding\_site 324..332  
FT /note- "Putative cation binding domain"  
FT Binding\_site 385..394



```

FT /note= "Putative cation binding domain"
FT Binding_site 441..449
FT /note= "Putative cation binding domain"
FT Domain 1040..1044
FT /label= Cytoplasmic domain
FT /note= "Conserved in virtually all integrins"
PN W09219647-A.
PD 12-NOV-1992.
PF 27-APR-1992; U03527.
PR 03-MAY-1991; US-695564.
PA (SRI ) SCRIPPS RES INST.
PI Quaranta V, Tamura RN;
DR N-PSDB; 031188.
DR Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used
FT for prodn. of antibodies and in detection of integrin sub-units
FT in body samples
PS Disclosure; Page 69-73; 115pp; English.
CC The sequences given in R28821-22 are the human alpha 6A and 6B
CC integrin subunits. Integrins are a family of cell surface receptors
CC which serve cellular adhesion functions. These receptors form a link
CC between the extracellular matrix and the cytoskeleton through their
CC binding to various extracellular components. Each integrin receptor
CC is a heterodimer comprised of an alpha and a beta subunit. Each alpha
CC subunit tends to associate with only one type of beta subunit but
CC there are several exceptions to this rule. These integrins correspond
CC to the laminin receptor. The cytoplasmic domain of the 6A and 6B
CC integrins differs from previously isolated alpha 6 integrins. The
CC human alpha 6B was isolated from human choriocarcinoma cell line JAR
CC by immunoprecipitation studies.
SQ Sequence 1073 AA;

Query Match 4.4%; Score 94; DB 5; Length 1073;
Best Local Similarity 25.3%; Pred. No. 2,18e+01;
Matches 20; Conservative 23; Mismatches 29; Indels 7; Gaps 7;

Db 997 vrvtfpskt-vaqsgvpgwllviallgmlallvflwkcgfk-rnk-kdhydnt 1053
QY 208 VAVSIYVGLSRVSDYK-HWMSDVLTLQGLAV-AILVAV-YVSDPFKERTSKERKEED 264
Db 1054 ykxaelhagpsdkerlts 1072
QY 265 SHIT-LHEPTTGNHYPN 282

RESULT 11
ID W20612 standard; protein; 276 AA.
AC W20612;
DE H. pylori protein.
KW Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh O, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR N-PSDB; T67865.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61; Page 1040-1041; 148pp; English.
CC The present sequence is a Helicobacter pylori protein.
CC H. pylori infection or to identify H. pylori polypeptide binding
CC compounds, useful as potential H. pylori life cycle activators or
CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
CC determined from overlapping contigs generated by mechanically

```

```

CC shearing the bacterial DNA. The sequences were analysed for ORF of
CC at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from
CC various ORF were analysed for significant homology to other known
CC or expected membrane proteins. Having identified and determined
CC the sequences of interest, particular regions can be isolated from
CC H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 276 AA;

Query Match 4.3%; Score 92; DB 21; Length 276;
Best Local Similarity 24.7%; Pred. No. 3,00e+01;
Matches 18; Conservative 21; Mismatches 28; Indels 6; Gaps 6;

Db 179 emlvsnppaiarnypleksvikephealfgvgkdellkeiv-flaakikipfl-vcemg 236
QY 83 NLLHSNS-FISNNY-IA-TIYKAIGTFLFGAA-ASQSLTDIAKYSIGRLRPHLDVCDPD 138
Db 237 ydqlkelkeclef 249
QY 139 WSKINGSDGYIEY 151

RESULT 12
ID R28822 standard; protein; 1091 AA.
AC R28822; 1993 (first entry)
DE Alpha 6B integrin subunit.
KW Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;
KW extracellular matrix; cytoskeleton; heterodimer; laminin receptor;
KW immunoprecipitation; JAR; choriocarcinoma.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 1012..1037
FT /label= Transmembrane domain
FT Peptide 1..23
FT /label= Signal peptide
FT Protein 24..1091
FT /label= Mature protein
FT Modified_site 223
FT /label= Glycosylation_site
FT Modified_site 284
FT /label= Glycosylation_site
FT Modified_site 370
FT /label= Glycosylation_site
FT Modified_site 513
FT /label= Glycosylation_site
FT Modified_site 731
FT /label= Glycosylation_site
FT Modified_site 748
FT /label= Glycosylation_site
FT Modified_site 891
FT /label= Glycosylation_site
FT Modified_site 927
FT /label= Glycosylation_site
FT Modified_site 958
FT /label= Glycosylation_site
FT Binding_site 230..238
FT /note= "Putative cation binding domain"
FT Binding_site 324..332
FT /note= "Putative cation binding domain"
FT Binding_site 386..394
FT /note= "Putative cation binding domain"
FT Binding_site 441..449
FT /note= "Putative cation binding domain"
FT Domain 1040..1044
FT /label= Cytoplasmic domain
FT /note= "Conserved in virtually all integrins"
FT Misc_difference 1044..1045
FT /note= "Position of deletion of alpha 6A"
PN W09219647-A.
PD 12-NOV-1992.

```



CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.  
SQ Sequence 170 AA;

Query Match 4.2%; Score 88; DB 21; Length 170;  
Best Local Similarity 31.8%; Pred. NO. 5.62e+01;  
Matches 21; Conservative 19; Mismatches 20; Indels 6; Gaps 6;

Db 41 qltstak-qilmv-iflglaifiwknldrvkeilmvialkeypmqlvaisvsnl-kei 97  
QY 197 RLLRPILQFGLVAVSIYGLSR-VSDYKHHSVDLTGLIOGALVAI-LVAVYSDFFKER 254

Db 98 ss-kek 102  
QY 255 TSFKER 260

Search completed: Tue Nov 4 10:24:46 1997  
Job time : 74 secs.

**This Page Blank (uspto)**





```

6
RESULT      S15308      #type complete
ENTRY      hypothetical protein 12.8 - Salmonella typhimurium
TITLE      #formal_name Salmonella typhimurium
ORGANISM   21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
DATE       10-Nov-1995
ACCESSIONS S15308
REFERENCE   JIANG, X.M.; Neal, B.; Santiago, F.; Lee, S.J.; Romana, L.K.;
#authors   Reeves, P.R.
#journal   Mol. Microbiol. (1991) 5:695-713
#title     Structure and sequence of the rfb (O antigen) gene cluster of
#cross-references M0ID:91360454
#accession S15308      preliminary
#status    #molecule_type DNA
#residues  1-430 #label MOL
SUMMARY    #length 430 #molecular-weight 48627 #checksum 6054

Query Match      5.0%; Score 105; DB 9; Length 430;
Best Local Similarity 31.1%; Pred. No. 7.82e-01;
Matches 19; Conservative 14; Mismatches 27; Indels 1; Gaps 1;

Db 178 YIYRGISVDKLSIVLPGVMISLCYIVYRIKIVHVT-TKSHYIAIRSSGFFI 236
QY 48 YPYHDTAASTVLLVGVGLPVSSIIIGETLSVVCNLLHNSFISNNYIATYIKAGTFL 107
Db 237 f 237
QY 108 F 108

7
RESULT      S28760      #type fragment
ENTRY      NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - blue
TITLE      #muscel mitochondrion (SGC4) (fragment)
ORGANISM   #formal_name mitochondrion Mytilus edulis #common_name blue
DATE       20-Feb-1995 #sequence_revision 26-Jul-1996 #text_change
REFERENCE   S28760
#authors   Hoffmann, R.J.; Boore, J.L.; Brown, W.M.
#journal   Genetics (1992) 131:397-412
#title     A novel mitochondrial genome organization for the blue
#accession S28760      muscel, Mytilus edulis.
#status    #molecule_type DNA
#residues  1-441 #label BRO
#cross-references EMBL:M83762

GENETICS
#gene      ND5
#genome    mitochondrion
#genetic_code SGC4
KEYWORDS   membrane-associated complex; mitochondrion; NAD; oxidative
#phosphorylation; oxidoreductase; respiratory chain
SUMMARY    #length 441 #checksum 3300

Query Match      5.0%; Score 106; DB 12; Length 441;
Best Local Similarity 28.8%; Pred. No. 6.06e-01;
Matches 23; Conservative 25; Mismatches 28; Indels 4; Gaps 4;

Db 330 LPVSMGAMSVAVLSGA-PFMSSGFFSKDLMIELSMDSVYGCYILEIIGI-IFTSFY 387
QY 15 LCVLLASMPMAVLKGIQYVFGFGFCCKNSINYPHDTAASTVLLVGVGLPVSSIIIL 74
Db 388 sari-vfsvmlgsn-yvnns 405
QY 75 GETLSVVCNLLHNSFISNN 94

8
RESULT      S28038      #type fragment
ENTRY      penicillin-binding protein 1a - Streptococcus pneumoniae
TITLE      #formal_name Streptococcus pneumoniae
ORGANISM   17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
DATE       17-Apr-1993
ACCESSIONS S28038
REFERENCE   MARTIN, C.; Sibold, C.; Hakenbeck, R.
#authors   ENBO J. (1992) 11:3831-3836
#journal   Relatedness of penicillin-binding protein la genes from
#title     different clones of penicillin-resistant Streptococcus
#cross-references M0ID:91360454
#accession S28038      translation not shown
#status    #molecule_type DNA
#residues  1-719 #label MAR
SUMMARY    #length 719 #checksum 2230

Query Match      5.0%; Score 105; DB 10; Length 719;
Best Local Similarity 28.6%; Pred. No. 7.82e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 408 VYNWDRGYF-GNITLQVALQGRNVPVETINKVGLNRKFTL-NGIGIDYPSIHYSNAI 465
QY 32 IYFQRGFCCKNSINYPHDTAASTVLLVGVGLPVSSIIIGETLSV-YCNLLHNSF 90
Db 466 ssn 468
QY 91 ISN 93

9
RESULT      A42893      #type complete
ENTRY      penicillin-binding protein 1A - Streptococcus pneumoniae
TITLE      #formal_name Streptococcus pneumoniae
ORGANISM   07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
DATE       07-Apr-1994
REFERENCE   A42893
#authors   Martin, C.; Briesse, T.; Hakenbeck, R.
#journal   J. Bacteriol. (1992) 174:4517-4523
#title     Nucleotide sequences of genes encoding penicillin-binding
#cross-references M0ID:91360454
#accession A42893      proteins from Streptococcus pneumoniae and Streptococcus
#status    #molecule_type DNA      oralis with high homology to Escherichia coli
#residues  1-719 #label MAR      penicillin-binding proteins 1A and 1B.
#cross-references GB:M90527
SUMMARY    #length 719 #molecular-weight 79701 #checksum 2518

Query Match      5.0%; Score 105; DB 10; Length 719;
Best Local Similarity 28.6%; Pred. No. 7.82e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 408 VYNWDRGYF-GNITLQVALQGRNVPVETINKVGLNRKFTL-NGIGIDYPSIHYSNAI 465
QY 32 IYFQRGFCCKNSINYPHDTAASTVLLVGVGLPVSSIIIGETLSV-YCNLLHNSF 90
Db 466 ssn 468
QY 91 ISN 93

10
RESULT      S28038      #type fragment
ENTRY      penicillin-binding protein 1a - Streptococcus pneumoniae
TITLE      #formal_name Streptococcus pneumoniae
ORGANISM   17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
DATE       17-Apr-1993

```

```

18-Nov-1994
S28038
#accessions S28031
#reference Martin, C.; Sibold, C.; Hakenbeck, R.
#journal EMBO J. (1992) 11:3831-3836
#title Relatedness of penicillin-binding protein la genes from
different clones of penicillin-resistant Streptococcus
pneumoniae isolated in South Africa and Spain.
#accession S28038
#status translation not shown
#molecule_type DNA
#residues 719
#cross-references EMBL:X67873
SUMMARY
#length 719 #checksum 2709

Query Match 5.0%; Score 105; DB 10; Length 719;
Best Local Similarity 28.6%; Pred. No. 7.82e-01;
Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;

Db 408 vynwgrgvf-gnltlqalqsrnpavetlnkvgnraktfi-ngldgdypsihysnai 465
QY :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
32 IYFQGFCDNSINYPHDSIAASTVLIIVGVLPVSVIIIGETLSV-YCNLLHSNSF 90

Db 466 ssn 468
QY ||
91 ISN 93

RESULT 11
ENTRY JCI173
#type complete
TITLE cycloheximide resistance protein - yeast (Candida maltosa)
ORGANISM #formal_name Candida maltosa
DATE 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change
30-Sep-1993

ACCESSIONS JCI173
REFERENCE JCI173
#authors Sasnauskas, K.; Jomantiene, R.; Lebediene, E.; Lebedys, J.;
Januska, A.; Janulaitis, A.
#journal Gene (1992) 116:105-108
#title Cloning and sequence analysis of a Candida maltosa gene which
confers resistance to cycloheximide.
#cross-references MUID:92331941
#accession JCI173
#molecule_type DNA
#residues 1-552 #label SAS
#cross-references GB:M64932

GENETICS
#gene CYHR
#keywords antibiotic resistance
SUMMARY
#length 552 #molecular-weight 61400 #checksum 9303

Query Match 4.9%; Score 104; DB 11; Length 552;
Best Local Similarity 25.7%; Pred. No. 1.01e+00;
Matches 18; Conservative 29; Mismatches 17; Indels 6; Gaps 5;

Db 450 ligmalfavagfifqtlfnymavsfkveylasvfesnafrfsvsagafplgralynnl 509
QY :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
62 LVGVGL-PVSSIIIGETLSVYCNLLHSNSFISNNYIA-TYIKAI--GTF-LFGAASQSL 116

Db 510 s-ldkfpqgw 518
QY :||||:
117 TDIAKYSIGR 126

RESULT 12
ENTRY S58012
#type fragment
TITLE probable olfactory receptor tpcr106 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change
19-Apr-1996

ACCESSIONS S58012
REFERENCE S57995
#authors Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.

```

```

submitted to the EMBL Data Library, July 1995
#description Male germ cells from several mammalian species express a
specific repertoire of olfactory receptor genes.
#accession S58012
#status preliminary
#molecule_type mRNA
#residues 1-157 #label VAN
#cross-references EMBL:X89667
SUMMARY
#length 157 #checksum 7167

Query Match 4.7%; Score 99; DB 13; Length 157;
Best Local Similarity 28.3%; Pred. No. 3.50e-00;
Matches 26; Conservative 26; Mismatches 31; Indels 9; Gaps 6;

Db 50 phff--cdlyalmKiscstytlnlmhtegavvisgalafit--asya--cillvvlri 103
QY :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
129 PHFLDVCDP-DWSKXNSCDGVIEYICRGNAERVKRGRLSFYSGHSFSMYCMLFVALYL 187

Db 104 ps-akgrvktfstcgshltvval-fygtlswv 133
QY :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
188 QARMKGDHARLLRPTLQFGLVAVSIYVGLSRV 219

RESULT 13
ENTRY I38164
#type complete
TITLE hypothetical protein 9 - Pseudomonas sp.
ORGANISM #formal_name Pseudomonas sp.
DATE 31-Jul-1992 #sequence_revision 31-Jul-1992 #text_change
18-Jun-1993

ACCESSIONS I38164
REFERENCE A38164
#authors Crouzet, J.; Levy-Schil, S.; Cameron, B.; Cauchois, L.;
Rigault, S.; Rouyez, M.C.; Blanche, F.; Debussche, L.;
Thibaut, D.
#journal J. Bacteriol. (1991) 173:6074-6087
#title Nucleotide sequence and genetic analysis of a
13.1-kilobase-pair Pseudomonas denitrificans DNA fragment
containing five cob genes and identification of structural
genes encoding cob(1)alamin adenosyltransferase, cobyrinic
acid synthase, and bifunctional cobinamide
kinase-cobinamide phosphate guanylyltransferase.
#cross-references MUID:92011366
#accession I38164
#status preliminary
#molecule_type DNA
#residues 1-261 #label CRO
#cross-references GB:M62856
SUMMARY
#length 261 #molecular-weight 27374 #checksum 8744

Query Match 4.7%; Score 100; DB 8; Length 261;
Best Local Similarity 31.1%; Pred. No. 2.74e-00;
Matches 19; Conservative 18; Mismatches 21; Indels 3; Gaps 3;

Db 102 vlkalpfilliaalyfglkpnmgdvdqh-srvtpfvftltlvp-ligfydgvfpgtgs 159
QY :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
198 LLRPTLQFGLVAVSIYVGL-SRVSDDKHHSVDYLTGLIOGALVAILVAVVYSDFFKERTS 256

Db 160 f 160
QY :
257 F 257

RESULT 14
ENTRY S68129
#type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 -
Protopterus dolloi mitochondrion (SGC1)
ORGANISM #formal_name mitochondrion Protopterus dolloi
DATE 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change
06-Dec-1996

ACCESSIONS S68129
REFERENCE S68128
#authors Zardoya, R.; Meyer, A.
#journal Genetics (1996) 142:1249-1263

```



#title	The complete nucleotide sequence of the mitochondrial genome of the lungfish (Protopterus dolloi) supports its phylogenetic position as a close relative of land vertebrates.									
#accession	S68129	preliminary								
##status	1-342	#label	ZAR							
##residues										
##cross-references	EMBL:L42813									
SUMMARY	#length	342	#molecular-weight	37535	#checksum	390				
Query Match		4.7%	Score	99;	DB	16;	Length	342;		
Best Local Similarity		25.0%;	Pred. No.	3.50e+00;						
Matches	22;	Conservative	32;	Mismatches	25;	Indels	9;	Gaps	8;	
Db	51	rateaatkylgaasim-llfism-inawagewmnlillspsa-tltlalaikig	107							
QY	160	RVEGRLSYSHSSFWTCMLFVALYIQARKMGDW--ARLRLTLOFLGIVAVSIYVGLS	217							
Db	108	-lap-mhfwlpevlvgvltmt-gailvt	132							
		::								
QY	218	RVS DYKHHW-SDVLTGLIQALVALVA	244							
RESULT	15									
ENTRY	B34284	#type	complete							
TITLE	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - sea urchin (Paracentrotus lividus) mitochondrion (SGC8)									
ORGANISM	#format_name Mitochondrion Paracentrotus lividus #common_name Common urchin									
DATE	29-Jun-1990	#sequence_revision	29-Jun-1990	#text_change						
	04-Nov-1994									
ACCESSIONS	B34284;	A26510								
REFERENCE	A34284									
#authors	Cantatore, P.; Roberti, M.; Rainaldi, G.; Gadaleta, M.N.; Saccone, C.									
#journal	J. Biol. Chem.	(1989)	264:10965-10975							
#title	The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of Paracentrotus lividus									
#cross-references	MUID:89291831									
#accession	B34284	#molecule_type	DNA							
		#residues	1-352	#label	CAN					
#cross-references	GB:J04815									
REFERENCE	A36510									
#authors	Cantatore, P.; Roberti, M.; Morisco, P.; Rainaldi, G.; Gadaleta, M.N.; Saccone, C.									
#journal	Gene	(1987)	53:41-54							
#title	A novel gene order in the Paracentrotus lividus mitochondrial genome.									
#cross-references	MUID:87248108									
#accession	A36510	#molecule_type	DNA							
		#residues	134-135,	'M',	137-148,	'K',	150-154,	'K',	156-183,	'M',
			185-189,	'M',	191-193;	202-211,	'K',	213-233,	'M',	225-230,
			232-238,	'M',	240,	'M',	242-247,	'M',	249-263	##label
										CAN
#cross-references	GB:M16516									
#note	the authors translated the nucleotide sequence using the standard genetic code; the authors translation is shown									
GENETICS										
#gene	ND2									
#genome	mitochondrion									
#genetic_code	SGC8									
CLASSIFICATION	#superfamily NADH dehydrogenase (ubiquinone) chain 2									
KEYWORDS	membrane-associated complex; mitochondrion; NAD; oxidoreductase									
SUMMARY	#length	352	#molecular-weight	39009	#checksum	8650				
Query Match		4.7%	Score	99;	DB	5;	Length	352;		
Best Local Similarity		32.6%;	Pred. No.	3.50e+00;						
Matches	28;	Conservative	22;	Mismatches	26;	Indels	10;	Gaps	9;	

**This Page Blank (uspto)**

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:20:41 1997; MasPar time 7.75 Seconds  
Tabular output not generated.  
Title: >US-08-842-827-4  
Description: (1-285) from US08842827.pep  
Perfect Score: 2119  
Sequence: 1 MFDKTRLPYVALDVLCVLLA.....HTTLHETPTGNHYPNSHQ 285  
Scoring table: PAM 150  
Gap 11  
Searched: 59021 seqs, 21210388 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot34  
Statistics: Mean 47.899; Variance 90.096; scale 0.532  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 454 21.4 341 11 YSX3\_CABEL HYPOTHETICAL 39.0 KD 7.00e-67  
2 121 5.7 345 7 NU2M\_XENLA NADH-UBIQUINONE OXIDO 6.87e-04  
3 105 5.0 430 8 RFBX\_SALTY RFEX PROTEIN 1.16e-01  
4 106 5.0 522 7 NU5M\_MATED NADH-UBIQUINONE OXIDO 8.52e-02  
5 105 5.0 719 7 PBPA\_STRPN PENICILLIN-BINDING PR 1.16e-01  
6 104 4.9 552 3 CYHR\_CANMA CYCLOHEXIMIDE RESISTA 1.57e-01  
7 101 4.8 345 11 YDE9\_SCHPO PUTATIVE MITOCHONDRIA 3.86e-01  
8 100 4.7 261 11 YCB3\_PSEDE HYPOTHETICAL 27.4 KD 5.19e-01  
9 99 4.7 332 7 NU2M\_PARLI NADH-UBIQUINONE OXIDO 6.96e-01  
10 100 4.7 3388 7 POLG\_DEN2P GENOME POLYPROTEIN (C 5.19e-01  
11 97 4.6 346 6 NU2M\_CHICK NADH-UBIQUINONE OXIDO 1.24e-00  
12 97 4.6 443 6 CDMF\_RHIME STRUCTURAL PROTEIN MD 1.24e-00  
13 98 4.6 676 2 NU2M\_COTJA CYTOCHROME C-TYPE BTO 9.31e-01  
14 98 4.6 704 7 PMA2\_LYCES PLASMA MEMBRANE ATPAS 1.24e-00  
15 97 4.6 723 2 CADA\_BACFI PROBABLE CADMIUM-TRAN 1.24e-00  
16 98 4.6 3391 7 POLG\_DEN26 GENOME POLYPROTEIN (C 9.31e-01  
17 98 4.6 3391 7 POLG\_DEN27 GENOME POLYPROTEIN (C 9.31e-01  
18 97 4.6 3391 7 POLG\_DEN28 GENOME POLYPROTEIN (C 1.24e-00  
19 95 4.5 161 11 YIAM\_HAEIN HYPOTHETICAL PROTEIN 2.20e-00  
20 95 4.5 346 6 NU2M\_COTJA NADH-UBIQUINONE OXIDO 2.20e-00  
21 96 4.5 347 7 NU2M\_PROVI NADH-UBIQUINONE OXIDO 1.66e-00  
22 96 4.5 347 7 NU2M\_HALGR NADH-UBIQUINONE OXIDO 1.66e-00

\*\*\*\*\*  
Release 2.1D John F. Collins, Blocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.  
\*\*\*\*\*

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:20:41 1997; MasPar time 7.75 Seconds  
Tabular output not generated.  
Title: >US-08-842-827-4  
Description: (1-285) from US08842827.pep  
Perfect Score: 2119  
Sequence: 1 MFDKTRLPYVALDVLCVLLA.....HTTLHETPTGNHYPNSHQ 285  
Scoring table: PAM 150  
Gap 11  
Searched: 59021 seqs, 21210388 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot34  
Statistics: Mean 47.899; Variance 90.096; scale 0.532  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES  
Result No. Score Query Match Length DB ID Description Pred. No.  
1 454 21.4 341 11 YSX3\_CABEL HYPOTHETICAL 39.0 KD 7.00e-67  
2 121 5.7 345 7 NU2M\_XENLA NADH-UBIQUINONE OXIDO 6.87e-04  
3 105 5.0 430 8 RFBX\_SALTY RFEX PROTEIN 1.16e-01  
4 106 5.0 522 7 NU5M\_MATED NADH-UBIQUINONE OXIDO 8.52e-02  
5 105 5.0 719 7 PBPA\_STRPN PENICILLIN-BINDING PR 1.16e-01  
6 104 4.9 552 3 CYHR\_CANMA CYCLOHEXIMIDE RESISTA 1.57e-01  
7 101 4.8 345 11 YDE9\_SCHPO PUTATIVE MITOCHONDRIA 3.86e-01  
8 100 4.7 261 11 YCB3\_PSEDE HYPOTHETICAL 27.4 KD 5.19e-01  
9 99 4.7 332 7 NU2M\_PARLI NADH-UBIQUINONE OXIDO 6.96e-01  
10 100 4.7 3388 7 POLG\_DEN2P GENOME POLYPROTEIN (C 5.19e-01  
11 97 4.6 346 6 NU2M\_CHICK NADH-UBIQUINONE OXIDO 1.24e-00  
12 97 4.6 443 6 CDMF\_RHIME STRUCTURAL PROTEIN MD 1.24e-00  
13 98 4.6 676 2 NU2M\_COTJA CYTOCHROME C-TYPE BTO 9.31e-01  
14 98 4.6 704 7 PMA2\_LYCES PLASMA MEMBRANE ATPAS 1.24e-00  
15 97 4.6 723 2 CADA\_BACFI PROBABLE CADMIUM-TRAN 1.24e-00  
16 98 4.6 3391 7 POLG\_DEN26 GENOME POLYPROTEIN (C 9.31e-01  
17 98 4.6 3391 7 POLG\_DEN27 GENOME POLYPROTEIN (C 9.31e-01  
18 97 4.6 3391 7 POLG\_DEN28 GENOME POLYPROTEIN (C 1.24e-00  
19 95 4.5 161 11 YIAM\_HAEIN HYPOTHETICAL PROTEIN 2.20e-00  
20 95 4.5 346 6 NU2M\_COTJA NADH-UBIQUINONE OXIDO 2.20e-00  
21 96 4.5 347 7 NU2M\_PROVI NADH-UBIQUINONE OXIDO 1.66e-00  
22 96 4.5 347 7 NU2M\_HALGR NADH-UBIQUINONE OXIDO 1.66e-00

23 95 4.5 366 2 BRB2\_MOUSE B2 BRADYKININ RECEPTO 2.20e+00  
24 96 4.5 397 7 P2X3\_RAT P2X PURINOCEPTOR 3 (A 1.66e+00  
25 520 10 VP7\_WTV NONSTRUCTURAL PROTEIN 1.66e+00  
26 96 4.5 615 6 NTNO\_BOVIN SODIUM-DEPENDENT NORA 2.20e+00  
27 95 4.5 727 2 CADA\_STAAU PROBABLE CADMIUM-TRAN 2.20e+00  
28 95 4.5 738 10 VU47\_HSV62 GLYCOPROTEIN U47 PREC 2.20e+00  
29 93 4.4 274 1 BACH\_HALSP HALORHODOPSIN PRECURS 3.85e+00  
30 94 4.4 418 9 SSR3\_HUMAN SOMATOSTATIN RECEPTOR 2.91e+00  
31 93 4.4 533 11 YADC\_SCHPO HYPOTHETICAL 62.2 KD 3.85e+00  
32 94 4.4 804 2 CADD\_STAAU PROBABLE CADMIUM-TRAN 2.91e+00  
33 94 4.4 947 7 PMA2\_ARATH PLASMA MEMBRANE ATPAS 2.91e+00  
34 94 4.4 1073 5 ITA6\_HUMAN INTEGRIN ALPHA-6 PREC 2.91e+00  
35 92 4.3 113 5 IMMO\_ECOLI IMMUNITY PROTEIN FOR 5.08e+00  
36 91 4.3 271 1 ATP6\_ECOLI ATP SYNTHASE A CHAIN 6.68e+00  
37 91 4.3 387 10 VP48\_NPVAC NITRATE/CHLORATE TRAN 6.68e+00  
38 91 4.3 590 2 CHU1\_ARATH PENICILLIN-BINDING PR 5.08e+00  
39 92 4.3 637 7 PBPA\_STROR NADH-UBIQUINONE OXIDO 6.68e+00  
40 91 4.3 652 7 NUSM\_PODAN HYPOTHETICAL 9.7 KD P 8.76e+00  
41 90 4.2 82 11 YICN\_LACLA HYPOTHETICAL 61.2 KD 8.76e+00  
42 90 4.2 532 11 YHB7\_YEAST FUSION GLYCOPROTEIN P 8.76e+00  
43 90 4.2 561 10 YGLF\_SV41 NADH-UBIQUINONE OXIDO 8.76e+00  
44 90 4.2 637 7 NUSM\_STRPU 180 KD SECRETORY PHOS 8.76e+00  
45 90 4.2 1463 7 PA2R\_BOVIN 180 KD SECRETORY PHOS 8.76e+00

ALIGNMENTS

RESULT 1  
ID YSX3\_CABEL STANDARD; PRT; 341 AA.  
AC Q10022;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 39.0 KD PROTEIN T28D9.3 IN CHROMOSOME II.  
GN T28D9.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACLOELOMATES; NEMATODA; SECERNITEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: WEAK, TO YEAST D9719.9.  
DR EMBL; U28738; G861267; -;  
DR WORMPEP; T28D9.3; CE02068.  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 71 91 POTENTIAL.  
FT TRANSMEM 122 142 POTENTIAL.  
FT TRANSMEM 223 243 POTENTIAL.  
FT TRANSMEM 257 277 POTENTIAL.  
SQ SEQUENCE 341 AA; 39028 MW; 89AE6E81 CRC32;

Query Match 21.4%; Score 454; DB 11; Length 341;  
Best Local Similarity 32.8%; Pred. No. 7.00e-67;  
Matches 82; Conservative 56; Mismatches 90; Indels 12; Gaps 11;  
Do 34 flataavtvtllgvsrgffcdoddairveyrkdttavqlmlyolvlvlnaatvlfvey 93  
Qy 18 LLASMPMAVLKLGQIYPPQFQFCNDNSINYPYHDSTAASTVLILVGLVPSVSIILGE- 76  
Db 94 yrmqvesnl-nprryrwnhhlvrlvrltyfgysqigfvmniaiivtkhvgvrlp 152  
Qy 77 -TL-SVYCNLLHNSNFI-SNNYIATYI-KATGTFLEGAA-SQSUT-DI-AKYSIGELRP 129  
Do 153 hfidvcklandtcvtgdshrytdtctgppelvlearksfysghsavlcyctwsaly1 212  
Qy 130 HFLDVCDDPDWSK-INC-SDGYIEYICRGAERVKRGLSFSYSGHSFMYCLFVALYL 187  
Do 213 qarlgpvlvnnrivvpsqltmiglsfsritdkkhwsdvlgvfigflavlyctcf 272  
Qy 188 QARKMGDWA-RLLRPLQGLVAVSIYVGLSRVSKHNSDVLTLGLQLGALVAILVAVY 246



[2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE; 92325042.  
 RA MARTIN C., BRIESE T., HAKENBECK R.;  
 RL J. BACTERIOL. 174:4517-4523(1992).  
 [3]  
 RN SEQUENCE OF 293-369 FROM N.A.  
 RP STRAIN=R6X;  
 RC MEDLINE; 95020625.  
 RA PEARCE B.J., YIN Y.B., MASURE H.R.;  
 RL MOL. MICROBIOL. 9:1037-1050(1993).  
 CC -1- FUNCTION: CELL WALL FORMATION.  
 CC -1- THESE STRAINS ARE PENICILLIN-SENSITIVE.  
 CC -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 DR EMBL; X67873; G47420; -.  
 DR EMBL; X67872; G47418; -.  
 DR PIR; S28038; G153768; -.  
 DR PIR; S28038; G28038.  
 KW PEPTIDOGLYCAN SYNTHESIS; ANTIBIOTIC RESISTANCE; CELL WALL;  
 KW MULTIFUNCTIONAL ENZYME.  
 FT ACT SITE 370 370 ACYLATED BY PENICILLIN (BY SIMILARITY).  
 FT DOMAIN 657 683 SER-RICH.  
 FT VARIANT 124 144 T -> A (IN STRAIN R6).  
 FT VARIANT 386 386 V -> I (IN STRAIN R6).  
 FT VARIANT 388 388 D -> E (IN STRAIN R6).  
 FT VARIANT 397 397 E -> K (IN STRAIN R6).  
 FT VARIANT 523 523 M -> I (IN STRAIN R6).  
 FT VARIANT 533 533 E -> D (IN STRAINS 63915 AND R6).  
 FT VARIANT 540 540 T -> S (IN STRAINS 63915 AND R6).  
 FT VARIANT 540 540 S -> N (IN STRAIN 63915).  
 FT VARIANT 657 657 S -> N (IN STRAIN 63915).  
 SQ SEQUENCE 719 AA; 79745 MW; 627E3A7B CRC32;  
 Query Match 5.0%; Score 105; DB 7; Length 719;  
 Best Local Similarity 28.6%; Pred. No. 1.16e-01;  
 Matches 18; Conservative 20; Mismatches 22; Indels 3; Gaps 3;  
 Db 408 vynydrqf-gnltqvalqgsnrvpavetlnkvlgnkrakf1-nglqidyvpslhyssnai 465  
 QY 32 IYPQRFEDKDSINYPHDSTAASVLLVGVLPVSSILGETLSV-YCNLLHSNSF 90  
 Db 466 ssn 468  
 QY 91 ISN 93  
 RESULT 6  
 ID CYHR CANVA STANDARD; PRT; 552 AA.  
 AC P32071;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE CYCLOHEXIMIDE RESISTANCE PROTEIN.  
 GN CYHR.  
 OS CANDIDA MALTOSEA (YEAST).  
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VSB-889;  
 RX MEDLINE; 92331941.  
 RA SASNAUSKAS K., JANUATIS A.;  
 RA JANUSKA S., JANUATIS A.;  
 RL GENE 116:105-108(1992).  
 CC -1- FUNCTION: PROBABLE TRANSPORTER. CONFERS RESISTANCE TO  
 CC CYCLOHEXIMIDE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
 CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY). BELONGS TO THE  
 CC CAR1/CYHR SUBFAMILY.  
 DR EMBL; M64932; G170849; -.  
 DR PIR; JC1173; JC1173.  
 KW ANTIBIOTIC RESISTANCE; TRANSPORT; TRANSMEMBRANE.  
 FT TRANSMEM 100 120 POTENTIAL.

FT TRANSMEM 137 157 POTENTIAL.  
 FT TRANSMEM 168 188 POTENTIAL.  
 FT TRANSMEM 194 213 POTENTIAL.  
 FT TRANSMEM 225 246 POTENTIAL.  
 FT TRANSMEM 262 282 POTENTIAL.  
 FT TRANSMEM 346 362 POTENTIAL.  
 FT TRANSMEM 381 399 POTENTIAL.  
 FT TRANSMEM 419 439 POTENTIAL.  
 FT TRANSMEM 445 464 POTENTIAL.  
 FT TRANSMEM 477 494 POTENTIAL.  
 FT TRANSMEM 518 539 POTENTIAL.  
 SQ SEQUENCE 552 AA; 61366 MW; AEP45552 CRC32;  
 Query Match 4.9%; Score 104; DB 3; Length 552;  
 Best Local Similarity 25.7%; Pred. No. 1.57e-01;  
 Matches 18; Conservative 29; Mismatches 17; Indels 6; Gaps 5;  
 Db 450 ligmalfavagilfqlfnyrnayvskfvyasvfnafsvagafplfgralynnl 509  
 QY 62 LVGVGL-PVSSILGETLSVYCNLLHSNSFISNNYA-TYIKAI--GTF-LFGAAASQSL 116  
 Db 510 s-ldkfpvgw 518  
 QY 117 TDIKYSIGR 126  
 RESULT 7  
 ID YDE9 SCHPO STANDARD; PRT; 345 AA.  
 AC Q10442;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PUTATIVE MITOCHONDRIAL CARRIER PROTEIN C12B10.09.  
 GN SPAC12B10.09.  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;  
 NL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (POTENTIAL).  
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.  
 DR EMBL; Z70721; E235727; -.  
 CR HYPOTHETICAL PROTEIN; MITOCHONDRION; INNER MEMBRANE; REPEAT;  
 KW TRANSMEMBRANE; TRANSPORT.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 83 103 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 220 240 POTENTIAL.  
 FT TRANSMEM 262 282 POTENTIAL.  
 FT TRANSMEM 296 316 POTENTIAL.  
 FT TRANSMEM 319 339 POTENTIAL.  
 SQ SEQUENCE 345 AA; 38451 MW; D059579A CRC32;  
 Query Match 4.8%; Score 101; DB 11; Length 345;  
 Best Local Similarity 28.7%; Pred. No. 3.86e-01;  
 Matches 29; Conservative 28; Mismatches 38; Indels 6; Gaps 5;  
 Db 93 glavdlslfpdtklkt--rlqakgfvknvgfghvrglsgilvgsapgaslf-fytten 149  
 QY 66 GLPVS-SIILGETLSVYCNLLHSNSFISNNYATYIKAI--GTF-LFGAAASQSLDIARY-S 123  
 Db 150 mkrlsqsglsgldqihmcasaslgiaacivrvptevikq 190  
 QY 124 I-GRLRPHFLVDCPDWMSKNSGDIYETICRGAERVK 163  
 RESULT 8  
 ID YCB9 PSEDE STANDARD; PRT; 261 AA.  
 AC P29942;



```

Db 1306 pfltsqqkdwipla-ltik-glnptaifltltsrsk-xrswp-lnealmavgmvs 1361
QY 184 ALYLQA-RMKGDWARLLRPTLQGLVAVSIYVG-LSRVSDYKHHWSDVLTGLIQLGALVAI 241
Db 1362 las 1364
QY 242 LVA 244

RESULT 11
ID NU2M_CHICK STANDARD; PRT; 346 AA.
AC P18937;
DT 01-NOV-1990 (REL. 15, CREATED)
DT 01-NOV-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 (EC 1.6.5.3).
GN ND2.
OS GALLUS GALLUS (CHICKEN).
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90230301.
RA DESJARDINS P., MORAIS R.;
RL J. MOL. BIOL. 212:599-634(1990).
CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
DR EMBL; X52392; G12962; -.
DR PIR; S10188; S10188.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 346 AA; 38333 MW; 60A7C67B CRC32;

Query Match 4.6%; Score 97; DB 6; Length 346;
Best Local Similarity 26.0%; Pred. No. 1.24e+00;
Matches 25; Conservative 33; Mismatches 28; Indels 10; Gaps 8;

Db 44 iskshhpraeatkyflqtgsaal-illfsm-tnaawtgqwdltqlnhptscsl-mltm 100
QY 153 ICRGNAERVKGRISFSGHSSFSMYCMLFVALYLQARMKGDW--ARLLRPTLQGLVAV 210
Db 101 ataiklqvp-f-hfwfpevlqgssltallstlm 134
QY 211 SIYGLSRVSDYKHHW-SDVLTG--LIQGLVAVLY 243

RESULT 12
ID MDML_YEAST STANDARD; PRT; 443 AA.
AC Q01846;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE STRUCTURAL PROTEIN MDML
GN MDML OR YML104C OR YH8339.15C.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92332595.
RA MCCONNELL S.J., YAPPE M.P.;
RL J. CELL BIOL. 118:385-395(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA SKELTON J., CHURCHER C.M., BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- FUNCTION: ESSENTIAL FOR MITOCHONDRIAL GROWTH. MEDIATES ORGANELLE
INHERITANCE.
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -|- SIMILARITY: LOW, TO VIMENTIN AND OTHER INTERMEDIATE FILAMENTS.
DR EMBL; X66371; G39167; -.
DR EMBL; 249210; G798896; -.
DR PIR; A42636; A42636.
DR LISTA; SC00593; MDML.

```

```

DR SGD; L0001051; MDML.
FT CONFLICT 88
FT CONFLICT 215
FT CONFLICT 379
FT CONFLICT 397
FT CONFLICT 412
SQ SEQUENCE 443 AA; 51470 MW; 621A08C9 CRC32;

Query Match 4.6%; Score 97; DB 6; Length 443;
Best Local Similarity 34.4%; Pred. No. 1.24e+00;
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Db 280 pfysqnsfsvkqicdlfslfainkanagwlr 311
QY 167 SFYSGHSSFSN-YCMLEVALYLQARMKGDWAR 197

RESULT 13
ID CCME_RHIME STANDARD; PRT; 676 AA.
AC P45404;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK.
GN CYCK.
OS RHIZORIUM MELILOTI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC RHIZOBIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AK631;
RA KERESZT A., SLASKA-KISS K., PUTNOKY P., BANFALVI Z., KONDOROSI A.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -|- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
CC -|- POSSIBLE SUBUNIT OF A HEME LYASE.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
POTENTIAL.
CC -|- SIMILARITY: BELONGS TO THE CCME/CYCK/CCCL1/NRFE/YCF5 FAMILY.
DR EMBL; X82560; G575372; -.
KW CYTOCHROME C-TYPE BIOGENESIS; TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 8
FT TRANSMEM 28
FT TRANSMEM 42
FT TRANSMEM 94
FT TRANSMEM 123
FT TRANSMEM 175
FT TRANSMEM 210
FT TRANSMEM 233
FT TRANSMEM 273
FT TRANSMEM 311
FT TRANSMEM 356
FT TRANSMEM 393
FT TRANSMEM 418
FT TRANSMEM 445
FT TRANSMEM 493
FT TRANSMEM 617
SQ SEQUENCE 676 AA; 72643 MW; C9947521 CRC32;

Query Match 4.6%; Score 98; DB 2; Length 676;
Best Local Similarity 35.3%; Pred. No. 9.31e-01;
Matches 18; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

Db 176 lylygvgsf-cfsfavaaliegidaawarwrvptlaawftlagiamg 225
QY 168 FYSGHSSFSMYCMLF-VALYLQARMKGDWARLLRP-TLQ-FGLVAVSIYVG 215

RESULT 14
ID PMA2_LYCES STANDARD; PRT; 704 AA.
AC P23980;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PLASMA MEMBRANE ATPASE 2 (EC 3.6.1.35) (PROTON PUMP) (FRAGMENT).

```





\*\*\*\*\*

MPSEARCH

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 11:57:37 1997; MasPar time 148.35 Seconds  
Tabular output not generated. 958.178 Million cell updates/sec

Title: >US-08-842-827-5  
Description: (1-1362) from US08842827.seq  
Perfect Score: 1362  
N.A. Sequence: 1 GGCCGACGCTGCAAAAGTT.....TTTAAAAAAA  
Comp: CCGGTCGAGACGTTTCAA.....AAATTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 9.450; Variance 6.196; scale 1.527

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
C 1	81	5.9	Q10572	Human Natriuretic Pep	4.20e-31
C 2	67	4.9	Q10572	Human Natriuretic Pep	6.44e-23
C 3	48	3.5	Q51746	Oligonucleotide probe	3.06e-12
C 4	45	3.3	Q51746	Oligonucleotide probe	1.27e-10
C 5	39	2.9	N81164	Base substituted E.co	1.83e-07
C 6	39	2.9	N81164	Base substituted E.co	1.83e-07
C 7	38	2.8	Q70467	Generic DNA sequence	6.00e-07
C 8	36	2.6	Q70467	Generic DNA sequence	6.24e-06
C 9	36	2.6	Q70468	Generic DNA sequence	6.24e-06
C 10	36	2.6	Q70468	Generic DNA sequence	6.24e-06
C 11	36	2.6	Q70466	Generic DNA sequence	6.24e-06
C 12	35	2.6	Q70469	Generic DNA sequence	1.99e-05
C 13	36	2.6	Q70468	Generic DNA sequence	6.24e-06
C 14	35	2.6	Q70469	Generic DNA sequence	1.99e-05
C 15	34	2.5	Q70470	Generic DNA sequence	6.25e-05

C 16	34	2.5	114 12	Q70465	Generic DNA sequence	6.25e-05
C 17	34	2.5	114 12	Q70466	Generic DNA sequence	6.25e-05
C 18	31	2.3	114 12	Q70470	Generic DNA sequence	1.82e-03
C 19	31	2.3	114 12	Q70470	Generic DNA sequence	1.82e-03
C 20	32	2.3	3871 2	N71302	HSV-1 gb and surround	5.98e-04
C 21	30	2.2	114 12	Q70471	Generic DNA sequence	5.45e-03
C 22	30	2.2	114 12	Q70471	Generic DNA sequence	5.45e-03
C 23	29	2.1	67 24	T14322	Primer used in the la	1.61e-02
C 24	29	2.1	114 12	Q70473	Generic DNA sequence	1.61e-02
C 25	29	2.1	114 12	Q70473	Generic DNA sequence	1.61e-02
C 26	28	2.1	114 12	Q70472	Generic DNA sequence	1.70e-02
C 27	27	2.0	81 21	T13611	DC43 TSAR library gen	1.35e-01
C 28	27	2.0	82 21	T13610	DC43 TSAR library gen	1.35e-01
C 29	27	2.0	565 6	Q35072	HCV envelope region n	1.35e-01
C 30	26	1.9	36 2	Q11195	Ballast Constituent c	3.80e-01
C 31	26	1.9	39 7	Q51787	Mixed oligonucleotide	3.80e-01
C 32	26	1.9	74 21	T13613	DC43 TSAR library gen	3.80e-01
C 33	26	1.9	75 21	T13612	DC43 TSAR library gen	3.80e-01
C 34	26	1.9	82 21	T13610	DC43 TSAR library gen	3.80e-01
C 35	26	1.9	565 6	Q35072	HCV envelope region n	3.80e-01
C 36	26	1.9	3871 2	N71302	HSV-1 gb and surround	3.80e-01
C 37	25	1.8	39 7	Q51787	Mixed oligonucleotide	1.05e+00
C 38	25	1.8	75 21	T13612	DC43 TSAR library gen	1.05e+00
C 39	24	1.8	201 2	N70195	Streptomyces protease	2.85e+00
C 40	24	1.8	498 3	N50034	Sequence encoding new	2.85e+00
C 41	24	1.8	501 3	N50024	Sequence encoding new	2.85e+00
C 42	24	1.8	501 3	N50032	Sequence encoding new	2.85e+00
C 43	24	1.8	501 3	N50030	Sequence encoding new	2.85e+00
C 44	24	1.8	787 28	T51183	Human breast specific	2.85e+00
C 45	25	1.8	1511 23	T29032	Endoglucanase (58 kDa	1.05e+00

ALIGNMENTS

RESULT 1  
ID Q10572 standard; DNA; 1047 BP.  
AC Q10572;  
DT 09-APR-1991 (first entry)  
DE Human Natriuretic Peptide Receptor B.  
KW NPBR; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;  
OS hyperaldosteronism; glaucoma; guanyl cyclase.  
PH Homo sapiens.  
FT Key Location/Qualifiers  
FT Peptide 1..22  
FT /label= signal sequence  
FT Protein 12  
FT /label= mature NPBR  
FT Domain 23..455  
FT /label= extracellular domain  
FT /note= "binds natriuretic peptides A, B and C]"  
FT Domain 456..456  
FT /label= transmembrane domain  
FT Domain 479..1047  
FT /label= cytoplasmic domain  
FT /note= "GC and protien kinase activity"  
FT Modified-site 24..26  
FT /label= N-glycos\_site  
FT Modified-site 35..37  
FT /label= N-glycos\_site  
FT Modified-site 161..163  
FT /label= N-glycos\_site  
FT Modified-site 195..197  
FT /label= N-glycos\_site  
FT Modified-site 244..246  
FT /label= N-glycos\_site  
FT Modified-site 277..279  
FT /label= N-glycos\_site  
FT Modified-site 349..351  
FT /label= N-glycos\_site  
FT Modified-site 600..602  
FT /label= N-glycos\_site  
PN WO9100292-A  
PD 10-JAN-1991.

```

PF 22-JUN-1990; U03586.
PR 23-JUN-1989; US-370673.
PA (GETH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI; 91-036711/05.
DR N-PSDB; Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1: 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prep'd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Query Match 5.9%; Score 81; DB 2; Length 1047;
Best Local Similarity 8.7%; Pred. No. 4.20e-31;
Matches 83; Conservative 265; Mismatches 598; Indels 9; Gaps 8;

Db 21 garntnnavnnhnsyavawrvgnvanavanngrannvdrnrvsnngacsnyna 80
Cp 1062 GAACATCACTGGGATGGTGTGGTGTGATACGGGAGACAGTCCCGTGAGAGGCCA 1003
Db 81 nnsavdknynhdnngngcvynaasvarnshvnnnnntagavasnksakndhyrtv 140
Cp 1002 T-CATGATCAAGGTGAATCGAGGAGGGCGGAGCAGCGGGCTCCTCGCCAAAGTGAAG 944
Db 141 rtgsankngnnvntghnwtarasnnyndartdrhntngnnvnnangsnsvnh 200
Cp 943 CGGGCTCGAGGTATAGACACAAATACAGATAGTATAGTATAGAGAGAGGCATGGGCCA 884
Db 201 nvyrnngnnnathnranrvnvcngnnmmhnnnnnnnnnnnnnnnnnnngvnyndvngn 260
Cp 883 GAGAGAAGGACTTCCTGGGCTTCTGGACTTGTGCTCATCCTCGATCTGATGTTTC 824
Db 261 snragntnatgrnwdrtnnannnnannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 320
Cp 823 TGAATGAGCTTCAGAGCAGTGAATCTGGCTGAATACAGAGTGGTTCAGACACTCAAGAG 764
Db 321 ngvnnngnnnnnagcnydngnnnyannnnnnnnnnnnnnnnnnngtrndgrnvnknmgryhgvtnv 380
Cp 763 TGAGGACGACGGCGCCCTATGGACACTTTGGCAATGT-CTGTGAAGAGACTGGCTGATGGC 705
Db 381 vmdkndrntdnvwngdndsg-dnnnaahysgranknnnwgrnrvnnwvkgansdnnnc 439
Cp 704 ACAGCAAGAGAGGAGAGAGCCACTTCCTATAGATGCTGCCAGTGGAGGCTTCTGAAT 645
Db 440 andndnsdktnnstnvanvngtntnmmgvssnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 499
Cp 644 CGTCGA-CCGGGACTTCTTCAGGTAATAGATCGCGTAGAATTCGCCCGTGTATGATCGCGA 586
Db 500 nngsnryhkgegrntnrgsyssnmbahgkynnnantghnkgvrvannkknkn 559
Cp 585 GGATGGCAATACCATCCACCGGCAC-AGAGCAGACGCTCATTTATGTTTCACCATTT 527
Db 560 ntrvnnnnkhmrdrnnnnhtrnnngacndnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 619
Cp 526 TTCAGTGGTACTTGTGCTCTCATCATTCAGTAAACCCCTCGGTGGTAAAGCTTGTATG 467
Db 620 rysnnndnvkgmannhnsnshgskssncvdrsnvknktdyngasnrtstannddna 679
Cp 466 GTGCTGTGCTGATGATGAGGAGGAGGGAGCGCCCATGAGAGGCACAAG-AGTTCGAG 408
Db 680 nyaknntannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 739
Cp 407 CGAGATGAGCAGCACCCTGTGCT--GCCGCTCCTCCTCGGTTGTTGAGCGCCGGG 350
Db 740 vrngnrvnrvnrdtrnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 799

```

Cp 349 CTGCGCGCGTCTTGTCTCTCCGGGAGCATCGCTTTGTGCTACTTGTAGTTTTCATGGCG 290  
 Db 800 nnnrnnnyannnnknvnnrtaynnnnkkanannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 859  
 Cp 289 CTGGCTCGCGCGAGCTCCGCGCGCGGAGAGCGTCCCGCAACAGCAGCAGCACACACC 230  
 Db 860 vtynsdnvgntansans-tmmvvtmnnndnyfctandndndvkvntngdaymrvvsgn 918  
 Cp 229 AGGCGCGCGGCTCGCTCTGCGGAGGCTGCTGCGGATAGTGGCGGCTGGGCCCGCGCT 170  
 Db 919 ngrngrnhanmanannndavssnrnrhrhndnrrnngvnhgncagvvgnk 973  
 Cp 169 CCGGCGCGCGGCTAGAGTGCAGCGGGGCTGCCTGCCCTCCAACTGCAGAAGT 115

RESULT 2  
 ID Q10572 standard; DNA; 1047 BP.  
 AC Q10572;  
 DT 09-APR-1991 (first entry)  
 DE Human Natriuretic Peptide Receptor B.  
 KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;  
 KW hyperaldosteronism; glaucoma; guanylyl cyclase.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /label= signal sequence  
 FT Protein 12  
 FT /label= mature NPRB  
 FT Domain 23..455  
 FT /label= extracellular domain  
 FT /note= "Binds natriuretic peptides A,B and C"  
 FT Domain 456..456  
 FT /label= transmembrane domain  
 FT Domain 479..1047  
 FT /label= cytoplasmic domain  
 FT /note= "GC and protein kinase activity"  
 FT Modified-site 24..26  
 FT /label= N-glycos-site  
 FT Modified-site 35..37  
 FT /label= N-glycos-site  
 FT Modified-site 161..163  
 FT /label= N-glycos-site  
 FT Modified-site 195..197  
 FT /label= N-glycos-site  
 FT Modified-site 244..246  
 FT /label= N-glycos-site  
 FT Modified-site 277..279  
 FT /label= N-glycos-site  
 FT Modified-site 349..351  
 FT /label= N-glycos-site  
 FT Modified-site 600..602  
 FT /label= N-glycos-site  
 PN WO9100292-A.  
 PD 10-JAN-1991.  
 PF 22-JUN-1990; U03586.  
 PR 23-JUN-1989; US-370673.  
 PA (GETH ) GENENTECH INC.  
 PI Chang M, Goeddel D, Lowe D;  
 DR WPI; 91-036711/05.  
 DR N-PSDB; Q10324.  
 PT Natriuretic protein receptor B - for diagnosis and treatment of  
 PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.  
 PS Claim 3; Fig 1: 49pp; English.  
 CC The sequence was derived from the DNA encoding natriuretic peptide  
 CC receptor B, NPRB, having guanylyl cyclase (GC) activity and protein  
 CC kinase activity. The DNA can be inserted into expression vectors  
 CC for the prodn. of the protein, opt. after being mutated to produce  
 CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=  
 CC 114,952). The protein (or variants) can be used in treatment of  
 CC natriuretic peptide disorders, and also to isolate peptides using  
 CC affinity chromatography. Antibodies with affinity for NPRB can  
 CC also be prep'd.

SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;  
 Query Match 4.9%; Score 67; DB 2; Length 1047;  
 Best Local Similarity 8.7%; Pred. No. 6.44e-23;  
 Matches 75; Conservative 246; Mismatches 552; Indels 13; Gaps 13;  
 Db 57 grannvdrnsvsnngnacsnyannasvdknyhdndnngvcvyna-asvarnaswh 115  
 QY 50 GGACCTTTAGAACGATTAGGTTG-TGACAGAGAAAGAGAGAGCGCGAGGAGGAGCA 108  
 Db 116 ronnntagavagsnakndhytntvrtgnsankngnnvtnhgnnnvtaraannndart 175  
 QY 109 AACACACACCTTCGTGAGTGGAGCA-GGCACCCCGCGCTGCTACTAGCGCGCGGCC 167  
 Db 176 ddrnhtntngvnnanngns-vnhvyrarngnnnnnathnrrangrnyvncgnnmh 234  
 QY 168 GGAGCGGGCGGACCGCCACTATCCGACGAGCGCTCGGCGAGGAGCGACCGCGCGC 227  
 Db 235 nnnnnanrntngdyvnyndvngnsragnrtatgrnwdnrdntrnnanannantvn 294  
 QY 228 CTGGG-TGTGTGCTGCTGTCGGGAGGTCTTCGGGGCGGAGGCTCGCGCGCAGC 386  
 Db 295 vntyrnn 354  
 QY 287 CAGCGCATGCAAACTACAGTACGACAAAGCGATCGTCCCGGAGAGCAAGAACGCGG 346  
 Db 355 ggrndgntnvmkngrryhgvt-gnvymdkndrdntdnvnmwngdndsdnnaahysg 413  
 QY 347 CAGCGCGGCTCAACAAACCCGAGGAGGAGCGGCGAGCGGCTGCTCATCTG 406  
 Db 414 anknnwtgrnnnwkgansdnncandndnsdctnnstnnanvngtntnnngvs 473  
 QY 407 CCTCGACCTTCTGCTCTTCATGCGCGGCTCCCTCTCATCATGAGACAGACAC 466  
 Db 474 snnnrknmmknnaasmvnrwnnnngnsnry-bkagrsrntnsrgssysnat-ah 531  
 QY 467 CATCAAGCCTTACCACCGAGGCTTTACTGCAATGATGAGCATCAAGTACCCTGAA 526  
 Db 532 gkyannaantghnkvvvaahkvnnkrrnnntnnnnnnkhmrdrvnnhtrnnagcndnn 591  
 QY 527 AACTGTGTGACAAATAAATGACGCTGTGCTGTGCGCGGAGTCTCATTCGCAATCCT 586  
 Db 592 n-cvntnycnrgsnnndnnndnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 650  
 QY 587 CGCATCATCAGCGGGAATCTACCGGATCTATTACTGAGAGTCCGGTCCGAGGAT 646  
 Db 651 vvdernvkntdygnasnrstannddnnnnanyaknnntnnnnnnnnnttgmnaadvsn 710  
 QY 647 TCAGAACCCCTACGTGCGAGCATCTATAAGCAAGTGGGCTGCTCTCTTGGCTGTC 706  
 Db 711 gnnnnnnanrsgnnyngndnsknvnrngnrnyrnnsdntnnnnnnnnnnnnnnrwa 770  
 QY 707 CATCAGCGATCTTTCACAGACATGCC-AAAGTGTCTATAGGCGGCTCGTCCT-CAC 764  
 Db 771 ndhanrndngnknrrnnknngtsndnnnnr-mnnyannnknvnnrntnnyannkx 829  
 QY 765 TTCTGTAGTGTCTGCAACCTGATTTACGCCAGATCAACTGCTCTGAAGGCTACATTCAG 824  
 Db 830 anannynnnhsvannkngrtnvnanandsvtynsdvngntansanstmmvntnn-d 888  
 QY 825 AACTACAGATGACAGGTGATGACAGCAAGTCCAGGAGCGAGCAAGTCTCTCTCTCT 884  
 Db 889 nycndannndndvykvtngdaymv 914  
 QY 885 GGCATGCCCTCTCTCCATGTACAC 910

## RESULT 3

ID Q51746 standard; cDNA; 91 BP.  
 AC Q51746;  
 DT 31-MAY-1994 (first entry)  
 DE Oligonucleotide probe MK14-A  
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;

KW ss.  
 OS Synthetic.  
 PN EP-571911-A.  
 PD 01-DEC-1993.  
 PF 24-MAY-1993; 108325.  
 PR 26-MAY-1992; US-889651.  
 PA (BECT ) BECTON DICKINSON CO.  
 PI Shank DD, Spears PA;  
 DR WPI; 93-378844/48.  
 PT New oligo:nucleotide probes specific for Mycobacteria - used for  
 detection and amplification of Mycobacteria nucleic acid in  
 samples  
 PS Claim 3; Page 14; 23pp; English.  
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
 CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
 CC cross reacted to a few non-mycobacterial spp. The probe may  
 CC be useful as an initial screen for mycobacterial infection.  
 CC See also Q51735-45 and Q51747-59.  
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.5%; Score 48; DB 9; Length 91;  
 Best Local Similarity 15.5%; Pred. No. 3.06e-12;  
 Matches 9; Conservative 44; Mismatches 5; Indels 0; Gaps 0;

Db 1 ggctccgcgcgssvshsvvvhvshhsvvvhvshvsvvvhvvhvvhvvhvhyv 58  
 Cp 287 GGCTGGCGGCGAGGCTCCGCCCGCGAGAGCTCCGCCAAGAGCAGCACACCC 230

## RESULT 4

ID Q51746 standard; cDNA; 91 BP.  
 AC Q51746;  
 DT 31-MAY-1994 (first entry)  
 DE Oligonucleotide probe MK14-A  
 DE Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
 KW ss.  
 OS Synthetic.  
 PN EP-571911-A.  
 PD 01-DEC-1993.  
 PF 24-MAY-1993; 108325.  
 PR 26-MAY-1992; US-889651.  
 PA (BECT ) BECTON DICKINSON CO.  
 PI Shank DD, Spears PA;  
 DR WPI; 93-378844/48.  
 PT New oligo:nucleotide probes specific for Mycobacteria - used for  
 detection and amplification of Mycobacteria nucleic acid in  
 samples  
 PS Claim 3; Page 14; 23pp; English.  
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
 CC (Q51735). It hybridized to all spp. of mycobacteria tested, but  
 CC cross reacted to a few non-mycobacterial spp. The probe may  
 CC be useful as an initial screen for mycobacterial infection.  
 CC See also Q51735-45 and Q51747-59.  
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.3%; Score 45; DB 9; Length 91;  
 Best Local Similarity 11.9%; Pred. No. 1.27e-10;  
 Matches 7; Conservative 45; Mismatches 7; Indels 0; Gaps 0;

Db 2 gctccgcgcgssvshsvvvhvshhsvvvhvshvsvvvhvvhvvhvvhvhyvsv 60  
 QY 139 GCCCGCGGTGCTACTAGCGCGCGCGCGAGCGCGGCGGCGGCGGCGGCGGCA 197

## RESULT 5

ID N81164 standard; DNA; 204 BP.  
 AC N81164;  
 DT 08-NOV-1990 (first entry)  
 DE Base substituted E.coli beta-galactosidase alpha-fragment.  
 DE E.coli beta galactosidase alpha-fragment; base substitutions; ss.  
 KW Escherichia coli.  
 FH Key Location/Qualifiers  
 FT misc\_feature 19..69

```

FT /*tag= a
FT /function-multiple cloning site
FT primer_bind 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80375.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.9%; Score 39; DB 1; Length 204;
Best Local Similarity 18.1%; Pred. No. 1.83e-07;
Matches 29; Conservative 68; Mismatches 61; Indels 2; Gaps 2;

Db 29 aycrcbcbgagcygcacbcyrragncycccggywccgagcygcaaycdchvgcgymr 88
QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
929 ATACCTGTGATGATGATGCTTACACGGAGCTGCTCGGTATCAGACCAACAGCAC 1046
Db 89 ttthyrzmrbnvdyndrsdaayccyrsvkydcynachdhdyvbybbvynhnnccbnhvhnh 148
QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
987 TTACCTGTGATGATGATGCTTACACGGAGCTGCTCGGTATCAGACCAACAGCAC 1046
Db 149 nancnccbnhvhnhvbnhnrwayrhdarddhvccv 188
QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1047 CATCCGATGATGCTGCGAGGATTTGCTCAGGAGCC 1086

RESULT 6
ID N81164 standard; DNA; 204 BP.
AC N81164;
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key misc_feature Location/Qualifiers
FT /*tag= a
FT /function-multiple cloning site
FT primer_bind 187..204
FT /*tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P: English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to

```

---

```

CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.9%; Score 39; DB 1; Length 204;
Best Local Similarity 16.8%; Pred. No. 1.83e-07;
Matches 25; Conservative 64; Mismatches 59; Indels 1; Gaps 1;

Db 47 yrraggnycccggywccgagcygcaaycdchvgcgymrttthyrzmrbnvdyndyn 106
Cp 576 TGAGATGCCCGACGCGACAGCAGCGCTCATTTATGTCACCCAGTTTTTCAGTGGGT 517
Db 107 rsdaaywccyrsvkydcynachdhdyvbybbvynhnnccbnhvhnhvbnhvb 166
Cp 516 ACTTGATGCTCT-CATCATTTGCAGTAAACCCCTCGGTGTAAGGCTTGATGCTGTTGTC 458
Db 167 nhnrwayrhdarddhvccvchccgat 195
Cp 457 TCGATGATGAGAGGAGGAGGCCGCCAT 429

RESULT 7
ID Q70467 standard; DNA; 114 BP.
AC Q70467;
DE Generic DNA sequence to generate a random TSAR peptide library.
TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key misc_feature Location/Qualifiers
FT /*tag= a
FT /note= *this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments).
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UNNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,

```





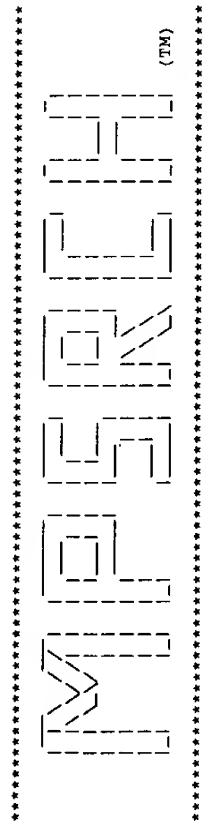
KW direct; rapid; detection; screening; treatment; generic; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 55..60  
 FT /\*tag= a  
 FT /note= "this sequence represents '2'; Z can be a  
 FT sequence of 6, 9 or 12 nucleotides (see  
 FT comments)"  
 FT WO9418318-A.  
 PN 18-AUG-1994.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U00977.  
 PR 01-FEB-1993; US-013416.  
 PR 30-DEC-1993; US-176500.  
 PR 31-JAN-1994; US-189331.  
 PA (UYN-) UNIV NORTH CAROLINA.  
 PI Fowlkes DM, Kay BK;  
 DR WPI: 94-279739/34.  
 DR P-SDB; R65154.  
 PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PT comprising a binding domain and an effector domain  
 PS Disclosure; Page 35; 25pp; English.  
 CC Q70469 is a generic DNA sequence used to generate random TSAR peptide  
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-  
 CC (TGC)(NNB)62(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction  
 CC sites (X is not the same as Y) that are not specified further. This  
 CC sequence generates peptides that are cloverleaf in structure. Other  
 CC generic sequences are shown in Q70465-68. Other specific peptides  
 CC generated by these generic sequences are shown in R65150-54. TSARs are  
 CC concatenated heterofunctional proteins or peptides, comprising at least  
 CC two functional regions - a binding domain with affinity for a ligand and  
 CC a second effector peptide portion that is chemically or biologically  
 CC active. They may further comprise a linker peptide between the 2 domains.  
 CC The oligonucleotides are also designed so that the expressed peptide  
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the  
 CC conformational rigidity to the peptides. These residues confer some degree of  
 CC a TSAR binding domain can be used in vivo to deliver a chemically or  
 CC biologically active moiety, eg. metal ion, radioisotope, peptide, toxin  
 CC or enzyme, to the specific target or on the cell. They can also replace  
 CC the function of macromolecules, eg. monoclonal or polyclonal antibodies  
 CC and therefore circumvent the need for complex methods of hybridoma  
 CC formation or in vivo antibody production. The TSARs are easily  
 CC characterised and have designed activity allowing direct and rapid  
 CC detection in a screening process.  
 CC Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;  
 SQ  
 Query Match 2.6%; Score 35; DB 12; Length 114;  
 Best Local Similarity 4.7%; Pred. No. 1.99e-05;  
 Matches 5; Conservative 32; Mismatches 70; Indels 0; Gaps 0;  
 Db 6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 65  
 QY 320 GATCGTCCGGAGACAGACGGCGGAGCCCGGCTCAACACACCCGAGGAGGAG 379  
 Db 65 btgcnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 112  
 QY 380 CGGCACAAAGCGGGTGTCTATCTGCTGCTGACCTCTCTGCTCT 426  
 RESULT 13  
 ID Q70468 standard; DNA; 114 BP.  
 AC Q70468;  
 DT 05-APR-1995 (first entry)  
 DE Generic DNA sequence to generate a random TSAR peptide library.  
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
 KW effector domain; concatenated heterofunctional protein; linker;  
 KW direct; rapid; detection; screening; treatment; generic; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 55..60  
 FT /\*tag= a  
 FT /note= "this sequence represents '2'; Z can be a  
 FT sequence of 6, 9 or 12 nucleotides (see

FT comments)"  
 FT WO9418318-A.  
 PN 18-AUG-1994.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U00977.  
 PR 01-FEB-1993; US-013416.  
 PR 30-DEC-1993; US-176500.  
 PR 31-JAN-1994; US-189331.  
 PA (UYN-) UNIV NORTH CAROLINA.  
 PI Fowlkes DM, Kay BK;  
 DR WPI: 94-279739/34.  
 DR P-SDB; R65154.  
 PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PT comprising a binding domain and an effector domain  
 PS Disclosure; Page 35; 25pp; English.  
 CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally  
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
 CC represented as follows: X(NNB)11(TGC)(NNB)62(NNB)10Y. X  
 CC and Y are flanking restriction sites (X is not the same as Y) that are  
 CC not specified further. Other generic sequences are shown in Q70466-68.  
 CC Other specific peptides generated by these generic sequences are shown in  
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,  
 CC comprising at least two functional regions - a binding domain with  
 CC affinity for a ligand and a second effector peptide portion that is  
 CC chemically or biologically active. They may further comprise a linker  
 CC peptide between the 2 domains. The oligonucleotides are also designed so  
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned  
 CC in, or flanking, the unpredicted or variant residues. These residues  
 CC confer some degree of conformational rigidity to the peptides. The TSARs  
 CC or comsps. comprising a TSAR binding domain can be used in vivo to  
 CC deliver a chemically or biologically active moiety, eg. metal ion,  
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the  
 CC cell. They can also replace the function of macromolecules, eg.  
 CC monoclonal or polyclonal antibodies and therefore circumvent the need  
 CC for complex methods of hybridoma formation or in vivo antibody  
 CC production. The TSARs are easily characterised in a screening process,  
 CC allowing direct and rapid detection in a screening process.  
 CC Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;  
 SQ  
 Query Match 2.6%; Score 36; DB 12; Length 114;  
 Best Local Similarity 3.6%; Pred. No. 6.24e-06;  
 Matches 4; Conservative 34; Mismatches 74; Indels 0; Gaps 0;  
 Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 62  
 Cp 389 CTTGTCGGCTCCCTCCCTCGGGTGTGTTGAGCGCGGCTCCCGCTTCCTGCTC 330  
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114  
 Cp 329 CGGAGCATCGCTTTGTCGTACTTGTAGTTTGCATGCGCTGCTGCGCG 278  
 RESULT 14  
 ID Q70469 standard; DNA; 114 BP.  
 AC Q70469;  
 DT 07-APR-1995 (first entry)  
 DE Generic DNA sequence to generate a random TSAR peptide library.  
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
 KW effector domain; concatenated heterofunctional protein; linker;  
 KW direct; rapid; detection; screening; treatment; generic; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 55..60  
 FT /\*tag= a  
 FT /note= "this sequence represents '2'; Z can be a  
 FT sequence of 6, 9 or 12 nucleotides (see  
 FT comments)"  
 FT WO9418318-A.  
 PN 18-AUG-1994.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U00977.  
 PR 01-FEB-1993; US-013416.  
 PR 30-DEC-1993; US-176500.  
 PR 31-JAN-1994; US-189331.









\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.  
\*\*\*\*\*  
MPsrch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 12:00:58 1997; MasPar time 719.25 Seconds  
Tabular output not generated. 972.045 Million cell updates/sec  
Title: >US-08-842-827-5  
Description: (1-1362) from US08842827.seq  
Perfect Score: 1362  
N.A. Sequence: 1 GCGCGAGCTCTGCAAAAGTT.....TTTAAAAAAAAAAAAAAA 1362  
Comp: CGCGCTGAGAGCTTTCAC.....AAATTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: EST-STS  
1: EST1 2: EST2 3: EST3 4: EST4 5: EST5 6: EST6 7: EST7 8: EST8  
9: EST9 10: EST10 11: EST11 12: EST12 13: EST13 14: EST14  
15: EST15 16: EST16 17: EST17 18: EST18 19: EST19 20: EST20  
21: EST21 22: EST22 23: EST23 24: EST24 25: EST25 26: EST26  
27: EST27 28: EST28 29: EST29 30: EST30 31: EST31 32: EST32  
33: EST33 34: EST34 35: EST35 36: EST36 37: EST37 38: EST38  
39: EST39 40: EST40 41: EST41 42: EST42 43: EST43 44: EST44  
45: EST45 46: EST46 47: EST47 48: EST48 49: EST49 50: EST50  
51: EST51 52: EST52 53: EST53 54: EST54 55: EST55 56: EST56  
57: EST57 58: EST58 59: EST59 60: EST60 61: EST61 62: EST62  
63: EST63 64: EST64 65: EST65 66: EST66 67: EST67 68: EST68  
69: EST69 70: EST70 71: EST71 72: EST72 73: EST73 74: EST74  
75: EST75 76: EST76 77: EST77 78: EST78 79: EST79 80: EST80  
81: EST81 82: EST82 83: EST83 84: EST84 85: EST85 86: EST86  
87: EST87 88: EST88 89: EST89 90: EST90 91: EST91 92: EST92  
93: EST93 94: EST94 95: EST95 96: EST96 97: EST97 98: EST98  
99: EST99  
EST-STS-TWO  
100: EST100 101: EST101 102: EST102 103: EST103 104: EST104  
105: EST105 106: EST106 107: EST107 108: EST108 109: EST109  
110: EST110 111: EST111 112: EST112 113: EST113 114: EST114  
115: EST115 116: EST116 117: EST117 118: EST118 119: EST119  
120: EST120 121: EST121 122: EST122 123: EST123 124: EST124  
125: EST125 126: EST126 127: EST127 128: EST128 129: EST129  
130: EST130 131: EST131 132: EST132 133: EST133 134: EST134  
135: EST135 136: EST136 137: EST137 138: EST138 139: EST139  
140: EST140 141: EST141 142: EST142 143: EST143 144: EST144  
145: EST145 146: EST146 147: EST147 148: EST148 149: EST149  
150: EST150 151: EST151 152: EST152 153: EST153 154: EST154  
155: EST155 156: EST156 157: EST157 158: EST158 159: EST159  
160: EST160 161: EST161 162: EST162 163: EST163 164: EST164  
165: EST165 166: EST166 167: EST167 168: EST168 169: EST169  
170: EST170 171: EST171 172: EST172 173: EST173 174: EST174

Statistics: Mean 11.544; Variance 2.649; scale 4.357

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description	Pred. No.
C 1	512	37.6	555	161	AA056693	2x81b05.sl Soares pre 0.00e+00
C 2	482	35.4	572	172	AA058383	2x81b05.sl Soares pre 0.00e+00
C 3	478	35.1	566	185	AA135741	z111f05.sl Soares pre 0.00e+00
C 4	440	32.3	449	173	W70040	z249b06.sl Soares fet 0.00e+00
C 5	405	29.7	451	83	H50533	yp08a11.sl Homo sapie 0.00e+00
C 6	402	29.5	480	54	H04639	yj49a08.sl Homo sapie 0.00e+00
C 7	401	29.4	428	191	AA156340	z054f04.sl Stratagene 0.00e+00
C 8	388	28.5	413	102	N68923	z269f12.sl Homo sapie 0.00e+00
C 9	380	27.9	402	103	N70674	z232a04.sl Homo sapie 0.00e+00
C 10	376	27.6	446	29	R63796	y155h04.sl Homo sapie 0.00e+00
C 11	357	26.2	431	29	R63741	y155h04.sl Homo sapie 0.00e+00
C 12	353	25.9	463	43	H02469	yj35d09.sl Homo sapie 0.00e+00
C 13	327	24.0	348	54	H04659	yj49a08.sl Homo sapie 0.00e+00
C 14	309	22.7	475	43	H02468	yj35d09.sl Homo sapie 0.00e+00
C 15	298	21.9	469	117	W31303	z264b01.sl Soares fet 0.00e+00
C 16	297	21.8	370	30	R71019	y150a11.sl Homo sapie 0.00e+00
C 17	291	21.4	442	60	H12563	yj2d810.sl Homo sapie 0.00e+00
C 18	287	21.1	422	69	H75437	yj96902.sl Homo sapie 0.00e+00
C 19	283	20.8	402	30	R71020	y150a11.sl Homo sapie 0.00e+00
C 20	273	20.0	395	10	T92854	ye27d03.sl Homo sapie 0.00e+00
C 21	269	19.8	279	69	H77957	ys10h08.sl Homo sapie 0.00e+00
C 22	269	19.8	398	138	AA048371	mj28d11.sl Soares mou 0.00e+00
C 23	267	19.6	485	60	H12562	yj2d10.sl Homo sapie 0.00e+00
C 24	259	19.0	275	64	H26102	y144e12.sl Homo sapie 0.00e+00
C 25	259	19.0	502	113	W00584	z269f12.sl Homo sapie 0.00e+00
C 26	253	18.6	292	20	R00691	ye74a10.sl Homo sapie 0.00e+00
C 27	254	18.6	344	69	H77856	ys09h08.sl Homo sapie 0.00e+00
C 28	249	18.3	332	40	R21805	y21g05.sl Homo sapie 0.00e+00
C 29	243	17.8	546	5	T71976	yc63g02.sl Homo sapie 0.00e+00
C 30	235	17.3	278	185	AA135784	z111f05.sl Soares pre 0.00e+00
C 31	227	16.7	241	40	R21755	y21g05.sl Homo sapie 0.00e+00
C 32	227	16.7	246	181	AA092191	l14679.seq F Fetal he 0.00e+00
C 33	228	16.7	246	122	W61641	m435a11.sl Life Tech 0.00e+00
C 34	223	16.4	225	109	HSC3MH032	H. sapiens partial cd 0.00e+00
C 35	217	15.9	223	110	HUM236B04B	Human aorta cDNA 5'-e 0.00e+00
C 36	207	15.2	292	20	R00690	ye74a10.sl Homo sapie 0.00e+00
C 37	204	15.0	263	145	R75377	MD05388 Mouse brain, 1.63e-298
C 38	202	14.8	203	165	C18274	Human placenta cDNA 5 6.13e-295
C 39	200	14.7	314	145	R75376	MD05388 Mouse brain, 2.29e-291
C 40	188	13.8	208	165	C16881	Human placenta cDNA 5 5.56e-270
C 41	159	11.7	372	39	R19424	y25b02.sl Homo sapie 1.10e-218
C 42	153	11.2	338	109	HSC3MH031	H. sapiens partial cd 3.78e-208
C 43	138	10.1	402	6	T78825	yc96f03.sl Homo sapie 5.69e-182
C 44	133	9.8	440	163	AA105463	mm92b06.sl Stratagene 2.68e-173
C 45	113	8.3	214	191	AA156375	z054f04.sl Stratagene 6.24e-139

ALIGNMENTS

RESULT 1  
LOCUS 1 AA056693 555 bp mRNA EST 17-SEP-1996  
DEFINITION 2x81b05.sl Soares pregnant uterus NBHPU Homo sapiens cDNA clone  
489201.3, similar to WP:T28D9.3 CE02068 ;  
ACCESSION AA056693  
NID 91549094  
KEYWORDS EST.

```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE
AUTHORS        Hillier, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M.,
                Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
                Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
                Travaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
                Wilson, R.
TITLE          The WashU-Merck EST Project
JOURNAL        Unpublished (1995)
COMMENT        Contact: Wilson RK
                WashU-Merck EST Project
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: est@watson.wustl.edu
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Seq primer: -40M13 fwd. from Amersham
                High quality sequence stop: 476.
                Location/Qualifiers
                1..555
                /organism="Homo sapiens"
                /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                oligo(dT) primer [5',
                AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'] ,
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                went through one round of normalization. Library
                constructed by M. Fatima Bonaldo."
                /clone="489201"
                /sex="female"
                /dev_stage="adult"
                /lab_host="DH10B"
                complement(<1..>555)

BASE COUNT      146 a 121 c 159 g 129 t

QUERY MATCH      37.6%; Score 512; DB 161; Length 555;
BEST LOCAL SIMILARITY 98.9%; Pred. No. 0.00e+00;
MATCHES 548; Conservative 0; Mismatches 0; Indels 6; Gaps 6;

Db 1 aaaaatcagtcgggcaaaagtcttcctacattctactgtctgtatgagattggagagca 60
Cp 1346 AAAAAATCATGCGGGCAAAAGTTTTTCCCTACTTCTACTGTGATGAGATTGGAGACA 1287
Db 61 gcaagaactgtgtcagcagcatttttacaacacagctcagagagtggtggcgaccta 120
Cp 1286 GCAAGAACCTGCTGTGCACAGTCATTTTACAAAACAGCTCAGAGGTGGGTGGCACCTA 1227
Db 121 catcatgttggtgattgttcctgtcaataatgtccacaggtgaagaattcttcctcg 180
Cp 1226 CATCATGTGTGTTGATGTTTCCCTGTCCTCAATGTCACAGGTGAAGGATTTCTCTCCG 1167
Db 181 gatagcagggcgagcgagcgagcgtcgtcttagctcttgaagagtcagacacgaagaa 240
Cp 1166 GATAGCAGGGCGAGCGAGGAGCGTCGCTTAGTCTTGAAGAGCTCAGACACGAAGAA 1107
Db 241 aactatgcagcagccaccaggggtcctttgagaaatccttcgcagaaacatcactgggatg 300
Cp 1106 AACTATGCAGCAGCCACCAGGGTCTCTTGAGCAAAATCTTCGACAGGATCACTGGGATG 1047
Db 301 gtgctgttggtctgatacgagacagtcctcggtgtagaagccatcatgatcaaggtgga 360
Cp 1046 GTGCTGTGTGTTGATGATCGCGAGACAGTCCCGTGTAGAGGCCATCATCATCAAGGTG-A 988

```

---

```

Db 361 acttcgagggggcgagcagggcggggtctctcccaagttaagcgggctcagagta 420
Cp 987 ACT-GCAGAGGGGGCGGAGCAGCGGGCTCTCCCAAGT-GAAGCGGGCTCAGGTA 930
Db 421 tagcaccaaaacacagcactagtgacatggagagagggcagtcagagagaagagactt 480
Cp 929 TAGCACCAAAATACAGCATAGTGTACATGGAGAAGGAGCGATGCCAGAGAGAGAGGACTT 870
Db 481 cctggctcctcggacttttctgtcctacacctctgcattctgaattcttgaatgta-cct 539
Cp 869 COTGGCTTCTCTGGACTTT-GCTGTCTCATCCTCTGCATCTGTA-GTTCTGAATGAGCCT 812
Db 540 tcagagcagttgat 553
Cp 811 TCAGAGCAGTTGAT 798

RESULT 2
LOCUS      AA058383 572 bp mRNA EST 19-SEP-1996
DEFINITION zK81b05.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
            489201 5' similar to TR:G1161100 G1161100 HYDROGEN
            PEROXIDE-INDUCIBLE PROTEIN ; .
ACCESSION  AA058383
NID        91551210
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 572)
AUTHORS    Hillier, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Travaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
            Wilson, R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT    Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 439.
            Location/Qualifiers
            1..572
            /organism="Homo sapiens"
            /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
            Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
            oligo(dT) primer [5',
            AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'] ,
            double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I
            and Eco RI sites of the modified pT7T3 vector. Library
            went through one round of normalization. Library
            constructed by M. Fatima Bonaldo."
            /clone="489201"
            /sex="female"
            /dev_stage="adult"
            /lab_host="DH10B"
            complement(<1..>572)

BASE COUNT      130 a 160 c 137 g 141 t 4 others

ORIGIN
Query Match      35.4%; Score 482; DB 172; Length 572;
Best Local Similarity 98.0%; Pred. No. 0.00e+00;
Matches 542; Conservative 0; Mismatches 1; Indels 10; Gaps 8;

```



```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS     Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 449)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE      Contact: Willson RK
JOURNAL    WashU-Merck EST Project
COMMENT    Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available royalty-free through LNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Insert Length: 1875 Std Error: 0.00
           Seq primer: mob REGA+ET
           High quality sequence stop: 227.
           Location/Qualifiers
           1..449
           /organism="Homo sapiens"
           /note="Organ: heart; Vector: p7T73D (Pharmacia) with a
           modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
           strand cDNA was primed with a Not I - oligo(dT) primer
           TGTACCAATCTCAAGTGGAGCGCGCCCACTTTTCTTTTCTTTT 3' ],
           double-stranded cDNA was size selected, ligated to Eco RI
           adapters (Pharmacia), digested with Not I and cloned into
           the Not I and Eco RI sites of a modified p7T73 vector
           (Pharmacia). Library went through one round of
           normalization to a Cot = 5. Library constructed by
           M.Fatima Bonaldo. This library was constructed from the
           same fetus as the fetal lung library, Soares fetal lung
           NbHL19W."
           /clone_lib="Soares fetal heart NbHL19W"
           /sex="unknown"
           /dev_stage="19 weeks"
           /lab_host="DH10B (ampicillin resistant)"
           complement(<1..>449)
           98 t
           130 c
           123 a
           98 c
           130 g
           98 t

BASE COUNT 123 a 98 c 130 g 98 t

ORIGIN
Query Match 32.3%; Score 440; DB 173; Length 449;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 447; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 1 taaaatcagtcgggcaaaagcttttccctacattctctgtctgtgatgagattgagagc 60
Cp 1347 TAAATAATCAGTCGGGCAAAAGCTTTTCCCTACATTTCTACGTCTGTGATGAGATTGGAGAGC 1288

Db 61 acgaagaacttgctgcagctcagctattttcaaaaacagctcaggagtggtggcgacct 120
Cp 1287 ACGAAGAAGCTTGCCTGTCAGCAGTCAATTTTCAAAAACAGCTCAGGAGGTGGTGCACCT 1228

Db 121 acatcatgttgggtgattgttctcgtcaataatagtcacacaggtgaaaggatttccttc 180
Cp 1227 ACATCATGTGTGGTGATGTGTTCTGTCATAAT-ATGCCACAGGTGAAGGATTTCCCTTC 1169

Db 181 cggatagcaggggagcggagggagcgtgctgtttagcttgaagggtcagacacgaag 240
Cp 1168 CGGATAGCAGGGGAGCGGAGGAGGCGTGTCTTAGCTTTGAAGAGGTGCAGACGGAAG 1109

Db 241 aaaaatgcagcagggccaccaggggctccttgagcaaatcctgccagaacatcactggga 300
Cp 1108 AAAAATATGCACAGGCCACCAGGGCTCCCTTGACAAATCCTGCCAGAACATCACTGGGA 1049

```

---

```

Db 301 tgggtgctgtggtctgtatcacgcagagacagtcacgtccgtgtatataagggccatcatgatcaaggtg 360
Cp 1048 TGGTGTGCTGTGGTCTGATACGCGGAGACAGTCCTCCGTGTAGAGGCCATCATGATCAAGGTG 989

Db 361 aactgcagagagggcgagcagcggtgctcctcccaagtgaagcggtgcttcagatgat 420
Cp 988 AACTGCGAGGAGGGCGGAGCAGCGGCGGCTCTCCCAAGTGAAGCGGCGCTGCGAGGTAT 929

Db 421 agccccaataacacgcatgtagtgcattatgga 449
Cp 928 AGCACCACAAATACAGCATAGTGTATATGGA 900

RESULT      5
LOCUS       H50533 451 bp mRNA EST 18-SEP-1995
DEFINITION YP08all.s1 Homo sapiens cDNA clone 186812 3'.
ACCESSION  H50533
NID        G990374
KEYWORDS   EST.
SOURCE     human clone-186812 primer=Promega -21ml3 library=Soares breast
           3NBH8st vector=p7T73D (Pharmacia) with a modified polylinker
           host=DH10B (ampicillin resistant) Rsitel=Not I Rsite2=Eco RI Adult
           human 1st strand cDNA was primed with a Not I - oligo(dT) primer
           [5' TGTACCAATCTCAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3' ],
           double-stranded cDNA was ligated to Eco RI adapters (Pharmacia),
           digested with Not I and cloned into the Not I and Eco RI sites of a
           modified p7T73 vector (Pharmacia). Library went through one round
           of normalization to a Cot = 20. Library constructed by Bento Soares
           and M.Fatima Bonaldo.
ORGANISM    Homo sapiens
Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 451)
AUTHORS     Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
           Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
           Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
           Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
           Wilson, R.
           The WashU-Merck EST Project
           Unpublished (1995)

TITLE      Contact: Willson RK
JOURNAL    WashU-Merck EST Project
COMMENT    Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           High quality sequence stops: 273
           Source: IMAGE Consortium, LNL
           This clone is available royalty-free through LNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Location/Qualifiers
           1..451
           /organism="Homo sapiens"
           /clone="186812"
           <1..>451
           120 a 98 c 135 g 94 t 4 others

BASE COUNT 120 a 98 c 135 g 94 t 4 others

ORIGIN
Query Match 29.7%; Score 405; DB 83; Length 451;
Best Local Similarity 97.8%; Pred. No. 0.00e+00;
Matches 442; Conservative 0; Mismatches 4; Indels 5; Gaps 6;

Db 1 ttctccatattctactgtctgtatgagattggagcagcagaactgtgttcagcag 60
Cp 1324 TTTCCTACATTTCTACTGTCTGTGATGATGGAGAGCAGCAAAAGACTTGTCTGT-CAGCAG 1266

Db 61 ttcatattcaaaaacagctcagaggtggtgggacacctacatcatgtttgtgtgattg 120
Cp 1265 T-CATTTTACAAAACAGCTCAGAGGTGGGTGG-CACCTACATCATGTTGTGTGTGATTG 1208

```

```

Db 121 ttctgtcaataatgtccacaggtgaaggatttcctccggtataggcagggcaggcca 180
|||||
Cp 1207 TTCTGTCAATAATGTCCACAGGTGAAGGATTTCCTTCGGATAG-CAGGGCAGG-CA 1150
|||||
Db 181 gggagagcgtctgttagtctttagttagtgagaggtcagacacgaagaatactatgcagcagcca 240
|||||
Cp 1149 GGGAGAGCGCTCGTCTTAGTCTTTGAAGAGGTGAGACACGAAGAATACTATGCAGCAGGCCA 1090
|||||
Db 241 ccagggtccttgagcaaatcctgcgaacatcactcaggtatggtctgttggtcgata 300
|||||
Cp 1089 CCAAGGCTCTTGAGCAATCTGCCAGACATCACTGGGATGGTGTGTGCTGTGATA 1030
|||||
Db 301 cgcgagacgtccctgttagaaggccatcatgatcaagggtgaactncaggagggcccgga 360
|||||
Cp 1029 CGCGAGACAGTCCCGTGTAGAAGGCCATCATGATCAAGTGAAGTGCAGGAGGGCCGGA 970
|||||
Db 361 gcaggggcctctgcgaagtgaaggcgggctcncaggtatagcaccataatcacgatag 420
|||||
Cp 969 GCAGCGGGCTCTCGCCAAAGTGAAGCGGGCTGCAGGTATAGCACCATAATACAGCATAG 910
|||||
Db 421 tgcacatnagaaggaggccatg-ccagagaag 451
|||||
Cp 909 TGTACATGGAGAGGAGGATGCCAGAGAAG 878
|||||

RESULT 6
LOCUS H04639 480 bp mRNA EST 20-JUN-1995
DEFINITION YJ48a08.s1 Homo sapiens cDNA clone 152054 3'.
ACCESSION H04639
NID 9867572
KEYWORDS EST.
SOURCE human clone=152054 library=Soares placenta NB2HP vector=pf73D
(Pharmacia) with a modified polylinker host=DH10B (ampicillin
resistant) primer=Promega -2ml3 Rsite1-Not I Rsite2-Eco RI Female
placenta obtained at birth (full term). 1st strand cDNA was primed
with a Not I - oligo(dT) primer [5].
AACGTGAAGAAATCCGCGCGAGGAATTTTTTTTTTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT7T3
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.
Homo sapiens
ORGANISM
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE The WashU-Merck EST Project
JOURNAL
COMMENT Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 270
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1..480
/organism="Homo sapiens"
/clone="152054"

BASE COUNT 121 a 107 c 133 g 112 t 7 others
ORIGIN

```

```

Query Match 29.5%; Score 402; DB 54; Length 480;
Best Local Similarity 97.7%; Pred. No. 0.00e+00;
Matches 422; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Db 1 ttctctacattctactctctgatgagattggagagcagcaaaacttctgtcagcagt 60
|||||
Cp 1324 TTTCCTCATATTCTACTCTCTGATGAGATTGGAGAGCAGCAAGAACTTGTCTCAGCACT 1265
|||||
Db 61 cattttcaaaaaacagctcagagagtggtggccacctacatcatgttgggtgattgttc 120
|||||
Cp 1264 CATTTTACAAAACAGCTCAGAGGTGGTGGCACCTTACATCATGTGTGGTGTATTGTC 1205
|||||
Db 121 ctgtcaataatgtccacaggtgaaggatttccttccggatagcagggcagcaggaggag 180
|||||
Cp 1204 CTGTCAATAATCTCCACAGGTGAAGGATTTCCTTCCGGATAGCAGGGCAGCAGGAG 1145
|||||
Db 181 acgctcgctcttagtcttgaagagctcagacacgaagaataactatgcagagccaccagg 240
|||||
Cp 1144 AGCTGCTCTTAGTCTTGAAGAGGTGACACGACGAGAAAACATATGCAGCAGGCCACGAG 1085
|||||
Db 241 gctccttgagcaaatcctgccagacatcatcctggtggtggtcttnggtctctgatacgga 300
|||||
Cp 1084 GCTCCTTGAGCAAAATCTGCCAGAACATCATCTGGATGTTGCTGTGCTGTATACGCA 1025
|||||
Db 301 gacagtcctgttagaagccatcatcaagtgaaactgaactgagagggcagcagcagg 360
|||||
Cp 1024 GACAGTCCCGTGTAGAAGGCCATCATGATCAAGTGAAGTGCAGGAGGGCCGCGAGCAGG 965
|||||
Db 361 cngggtcctt-gccaagtgaaagctgcctncagggtat-gcaccataatcacagcatagt-tac 417
|||||
Cp 964 CGGGCTCCTCGCCAAAGTGAAGCGGCTGCAGGTATAGCACCATAATACAGCATAGTGTAC 905
|||||
Db 418 atggagagaggag 429
|||||
Cp 904 ATGGAGAAGGAG 893
|||||

RESULT 7
LOCUS AA156340 428 bp mRNA EST 11-DEC-1996
DEFINITION 2054f04.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
590719 3'.
ACCESSION AA156340
NID 91727974
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 428)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)

TITLE The WashU-Merck EST Project
JOURNAL
COMMENT Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 329.
Location/Qualifiers
1..428
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1.
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:

```

```

Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Mini-ZAP XR Vector: -5', adaptor
sequence: 5', GAATTCGCGACGAG 3' -3', adaptor sequence: 5'
CTCGAGTCTTCTTTTTTTTTTTTTT 3'
/cclone_590719"
/cclone_lib="Stratagene preareas (#937208)"
/lab_hosts="SOLR cells (kanamycin resistant)"
complement(<cl_>428)

```

BASE COUNT	complement(<1..>428)		complement(<1..>428)	
ORIGIN	113 a	95 c	93 t	2 others
mrna				

Query Match	29.4%	Score 401;	DB 191;	Length 428;
Best Local Similarity	98.8%	Pred. No. 0.00e+00;		
Matches 419;	Conservative	0;	Mismatches 2;	Indels 3; Gaps 3;

Db 1 aaatcagtcgggcaaaagttttccctacattctactgtctgatgagattggagagcag 60

Cp 1345 AAAATCAGTCGGGAAAAGTTTTCCCTACATCTACTGCTGATGAGATTGGAGAGCAG 1286

Db 61 caagaacttgctgcagcagtcattttacaaaaaacagctcaggaggtgggtggcacctac 120

CP 1285 CAAGAACCTTGCTGTCAGCAGTCATTTTACAAAAACAGCTCAGGAGGTGGGTGGCACCCTAC 1226

DB 121 atcatgttggtgggaattgttcctgtcaataaagtccacagggtgaaggatttccttcctcg 180

CP 1225 ATCATGTTGGGAGATGGTTCCTGTCTATATATGTCACAGGTGAAGGATTTCCTTCGGG TTGG

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
1995	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024

b6  
b7C [REDACTED] 200

Cp 1105 ACTATGCAGCGGGCCACCGGGGCTCCTTGAGCAATCTCCAGAACATCATGGGATGG 1046

Db 301 tgccttggtcctatatcgcgcgagacagtcctccctgtagaagagcccatcatgcatcaagctgaa 360

Cp 1045 TGCTTGTGGTC-TGATACGGGAGACAGTCCCCTGTAGAAGGCCATCATGATCAAGGTGA 987

**Db** 361 ctgcangaggggcccaggacaggcgggtctctctctcgccaagtgaagcgggncttgaggat 420

986 CTGCAGGAGGGCGGAGCAGCGGGCTCCTGCCAAGT-GAAGCGGGCCT-GCAGGTAT 929

Db 421 agca 424

CP 928 AGCA 925

## RESULT 8

LOCUS	NC00923	413 bp	EST	13-MAR-1990
DEFINITION	2a69f12.s1	Homo sapiens	cDNA clone 297839	3'

NID 91225084  
 91225084  
 91225084

**SOURCE**  
human clone=297839 primer-m13 -40 forward library=Soares fetal  
NHU119W vector=pTZ19 (Pharmacia) with a modified polylinker

host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI

primer [5'-TGTTACCAATCTGAAGTGGGACGGCCGCATT-3']

RI sites of a modified pT7T3 vector (Pharmacia). Library went (Pharmacia), digested with Not I and cloned into the Not I and

through one round of normalization to a COL = 3. Library constructed by Bento Soares and M. Fatima Bonaldo.

REFERENCE : (bureau 1 to 413)

**AUTHORS**  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le M., Lennon, C., Marra, M.

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and

TITLE	WILSON, R. The WashU-Merck EST Project
-------	---

Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 402)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 High quality sequence stops: 97  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
 source  
 1..402  
 /organism="Homo sapiens"  
 /clone="294222"  
 <1..>402  
 106 a 88 c 109 g 91 t 8 others  
 Query Match 27.9%; Score 380; DB 103; Length 402;  
 Best Local Similarity 97.5%; Pred. No. 0.00e+00;  
 Matches 392; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

Db 1 ttaaaatcagtcggcgaagtttttccctacattctactgtctgatgagtgagag 60  
 Cp 1348 TTAANAATCAGTCGGCAAAAGTTTTTCCTACATCTACTGTGTGAGATGGAGAG 1289  
 Db 61 cagcaagaactnctgtcagcagtcattttacaaaacagctcangaggtgggtggcacc 120  
 Cp 1288 CAGCAAGAATCTGTCTGACGAGTCATTTTACAAAACAGCTCAGGAGGTGGGTGGCACC 1229  
 Db 121 tacatcagtgttggtgattgttcctgtcaataatntcacagtgaaagatttccttc 180  
 Cp 1228 TACATCAIGTGTGGTATGTTCCCTGTCATTAATGTCCACAGGTGAAGAGATTTCCTTC 1169  
 Db 181 cggatagcagggcagggcagggagagcgtcgtcttagtcttgaanaggtcagacacgaag 240  
 Cp 1168 CGGATAGCAGGGCAGCAGGAGAGAGCGTCTGTAGTCTTGAAGAGGTCAGACACGAAG 1109  
 Db 241 aaactatgcagcagccaccagggctccttgagcaaatcctgcagaaacatacctngga 300  
 Cp 1108 AANAATATGACGAGGACCACAGGGCTCTTGAGCAAAATCTGCCAGAACATCAGTGGGA 1049  
 Db 301 tgggtgtgtgtgtctgtatcacgcnacagtcctcctntagaagggccatcatgacaggt 360  
 Cp 1048 TGGTGTCTGT-TGGTCTGATACGGCAGACAGATCCCGTGTAGAAGGCCATCATGATCAAGGT 990  
 Db 361 gaactcagagggggccgancagcaggtgggtcctcgcaag 402  
 Cp 989 GAACCTGACGAGGGGGCCGGAGAGGC-GGGCTCCTCGCCAAG 949

RESULT 10  
 LOCUS R63796 446 bp mRNA EST  
 DEFINITION Y115H04.r1 Homo sapiens CDNA clone 139351 5'.  
 ACCESSION R63796  
 NID 9835675  
 KEYWORDS  
 SOURCE human clone-199351 library-Soares placenta N02HP vector-pt7r3D  
 (Pharmacia) with a modified polylinker host-DH10B (ampicillin  
 resistant) primer-M13Rpi Rsite1-Not I Rsite2-Eco RI Female placenta  
 obtained at birth (full term). 1st strand cDNA was primed with a

Not 1 - oligo(dT) primer [5'  
 AACGGAAGAAATTCGGCGCGAGGAATTTTTTTTTTTTTTTTTTTT 3'], double-stranded  
 cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the modified p7T3  
 vector. Library went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo.  
 Homo sapiens  
 Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;  
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;  
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;  
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 446)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 High quality sequence stops: 153  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
 source  
 1..446  
 /organism="Homo sapiens"  
 /clone="139351"

BASE COUNT 84 a 122 c 110 g 124 t 6 others  
 ORIGIN

Query Match 27.6%; Score 376; DB 29; Length 446;  
 Best Local Similarity 94.8%; Pred. No. 0.00e+00;  
 Matches 423; Conservative 0; Mismatches 17; Indels 6; Gaps 6;  
 Db 2 agcaagtgggtgcttctcttcttggctgctcagccagctcttccacagacattgcca 61  
 Qy 676 AGCAAGTGGGCTCTCTCTTTGGCTGTGGCATCAGCCAGCTTTTCACAGACATTGCCA 735  
 Db 62 aagtgtccatagggcgctcgtctcctcactcttctgagtgctgcaacctgatttcagcc 121  
 Qy 736 AAGTGTCCATAGGCGCCTGCGTCTCTACTTCTTGAAGTGTGCAACCTGATTTCAGCC 795  
 Db 122 agatcaactgctctgaaggtcatctcagaactacagatgcagaggtgtagacagaag 181  
 Qy 796 AGATCAACTGCTCTGAAGGCTACATTCAGAACTACAGATGCAGAGGTGATGACAGCAAAG 855  
 Db 182 tccaggnagccaggnagtccttctctgcccagctcttctctccatctacactatgc 241  
 Qy 856 TCCAGGAGCCAGGAAGTCTTCTCTCTGCCCAGTCCCTCTCTCCATGTACACTATGC 915  
 Db 242 tgtatttggtatatactgcagggccctctctcatttggggcgagagacc-gcctgctccgg 300  
 Qy 916 TGTATTGTGTATACCTGTCAGGGCCCTTCACTTGG-CGAGGAGCCCGCTGCTCCGG 974  
 Db 301 cccctctcgtcagttcaccttgatcatgatggcctctctacacgggagctgttngtatctc 360  
 Qy 975 CCCCTCTCTGAGTTCACTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1033  
 Db 361 agaccacagcagcaccattccagtgatgtctcggcagagatttgttcagggagcggcgggt 420  
 Qy 1034 AGACCAACAAGCAACCAT-CCAGTGTATGTTCTGG-CAGSATTTGCTCAAGG-AGCCCTGTGT 1090  
 Db 421 nggctgtcgtcagttttttttttgtgt 446  
 Qy 1091 GGCCTGCTGCATAGTTTCTTCTGCT 1116



```

RESULT 11 R63741 431 bp mRNA EST 26-MAY-1995
LOCUS Y115h04.s1 Homo sapiens cDNA clone 199351 3'
DEFINITION R63741
ACCESSION 9835620
NID EST
KEYWORDS human clone-139351 library-Soares placenta Nb2HP vector-pt7T3D
SOURCE (Pharmacia) with a modified polylinker host-DH10B (ampicillin
resistant) primer-promega -2lm13 Rsite1-Not I Rsite2-Eco RI Female
placenta obtained at birth (full term). 1st strand cDNA was primed
with a Not I - oligo(dT) primer [5'
AAGTGAAGAAATTCGGCGCCGAGGAATTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT7T3
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 100
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1..431
/organism="Homo sapiens"
/clone="139351"

BASE COUNT 111 a 91 c 123 g 97 t 9 others
ORIGIN

Query Match 26.2%; Score 357; DB 29; Length 431;
Best Local Similarity 96.0%; Pred. No. 0.00e+00;
Matches 406; Conservative 0; Mismatches 9; Indels 8; Gaps 8;

Db 1 aaatcagtcgggcaaaagtgtttccctacattctctgtctgtatgagattggagagcag 60
Cp 1345 AAAATCAGTCGGGCAAAAGTGTTCCTTACATTTCTACTGCTCATGATTCGAGAGCAG 1286
Db 61 caagaactgtgtcagcagcattttacaaaaaacagctcagagggtgggtggcactac 120
Cp 1285 CAAGAACTTGTCTGTCAGCAGTCATTTTACAAAAACAGCTCAGAGGTGGTGGCACCTAC 1226
Db 121 atcatgttggtgtgttctctgtcctaataatgtccacaggtgaaggatttccctccgg 180
Cp 1225 ATCATGTTGTGTGATTGTTCTCTCAATATGTCACAGCTGAAGGATTTCTTCGG 1166
Db 181 atagcagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 240
Cp 1165 ATAGCAGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1106
Db 241 actatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 300
Cp 1105 ACTATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1046

```

---

```

Db 301 gtgcttggtggtctnatacgcagagagcagtcocgntnaggaaagccatcctcagcaggt 360
Cp 1045 -TGCCTGTGG-TCTGATACGCGAGACAGTCCCGCTGTAG-AAGCCAT-CATGATCAAGGT 990
Db 361 gaacttcacagggagggccngagcagcagcagcagcagcagcagcagcagcagcagcagc 420
Cp 989 GAACCT-GCAGG-AGGGGCCGAGCAGCAGCGGGCTCTT-CGCCAAGT-GAAGCGGCGCTGCA 934
Db 421 ggt 423
Cp 933 GGT 931

RESULT 12 H02469 463 bp mRNA EST 20-JUN-1995
LOCUS YJ35d09.s1 Homo sapiens cDNA clone 150737 3'
DEFINITION H02469
ACCESSION 985402
NID EST
KEYWORDS human clone-150737 library-Soares placenta Nb2HP vector-pt7T3D
SOURCE (Pharmacia) with a modified polylinker host-DH10B (ampicillin
resistant) primer-promega -2lm13 Rsite1-Not I Rsite2-Eco RI Female
placenta obtained at birth (full term). 1st strand cDNA was primed
with a Not I - oligo(dT) primer [5'
AATCGAAGAAATTCGGCGCCGAGGAATTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the modified pT7T3
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 463)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 358
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1..463
/organism="Homo sapiens"
/clone="150737"

BASE COUNT 122 a 96 c 134 g 103 t 8 others
ORIGIN

Query Match 25.9%; Score 353; DB 43; Length 463;
Best Local Similarity 94.9%; Pred. No. 0.00e+00;
Matches 429; Conservative 0; Mismatches 11; Indels 12; Gaps 11;

Db 1 aaatcagctntgttaaaagtgtttccctacattctctgtatgagattggagagcag 60
Cp 1345 AAAATCAGTCGGGCAAAAGTGTTCCTTACATTTCTACTGCTGATGATTCGAGAGCAG 1286
Db 61 caagaactgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 120
Cp 1285 CAAGAACTTGTCTGTCAGCAGTCATTTTACAAAAACAGCTCAGAGGTGGTGGCACCTAC 1226

```



Db	121	atcatgttgggtgattgttctctgcaataatgtccacaggtagaaggatttccttcgg	180
Cp	1225	ATCATGTTGGTGATTTGCTTCTGCTCAATATGTCCACAGGTCAAAAGGATTTCTCTCCGG	1166
Db	181	atagcaggggcagcagcagcagcgtcgtcttagtttgaagaggtcagacacgaagaa	240
Cp	1165	ATAGCAGGGGCAGGCAGGAGGAGCGTGTCTTAGTTGAAGAGGTGACACACGAAGAAA	1106
Db	241	actatgcagcagggccaccagggctccttgagcgaatacttgcagaaacatcactg	300
Cp	1105	ACTATGCACAGCAGGCACCCAGGGCTCTTGAGCAAAATCTGCCAGAAACATCACTGGG	1047
Db	301	ggtgcttatgggtctgatacggcagacagctccggtgtaggaagccattatgataaa	360
Cp	1046	G-TGCTTGTGG-TCTGATACG-CGAGACAGTCCGCTGTAG-AAGGCCAT-CATGATCAA	993
Db	361	gggtgaactgcagcaggggcnagcagcgggcttctctgccaagtgaacgcgacctn	420
Cp	992	-GGTGAACATGCAGGAGGGCGGAGCAGCGGGCT-CCTGCCCAAGT-GAAGCGGCGCTG	936
Db	421	ccaggttaggcacnaatacagctagtnta	452
Cp	935	C-AGGTATAG-CACCAATACAGCATAGTGA	906
RESULT 13			
LOCUS	H04659	348 bp	mRNA EST 20-JUN-1995
DEFINITION	Y149a08.r1 Homo sapiens CDNA clone 152054 5'		
ACCESSION	H04659		
NID	9867592		
KEYWORDS	EST.		
SOURCE	human clone=152054 library=Soares placenta Nb2HP vector-pt7r3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13Rev Rsfcl-Not I Rsite2-Eco RI female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAGAAATTCGGCGCGGAGGAGTATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo.		
ORGANISM	Homo sapiens Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 348)		
AUTHORS	Hollier,L., Clark,N., Dubuca,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 257 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. LOCATION/Qualifiers 1..348 /organism="Homo sapiens" /clone="152054"		
FEATURES	source		
BASE COUNT	74 a	102 c	79 g 90 t 3 others
ORIGIN			

Query Match	24.0%	Score 327;	DB 54;	Length 348;
Best Local Similarity	98.0%;	Pred. No. 0.00e-00;		
Matches 341;	Conservative 0;	Mismatches 5;	Indels 2;	Gaps 2;
Dbb	1	gccagatacaactgctcgaaggctacattcagaactacagatgcagaggatgacacanca	50	
QY	793	GCCAGATCAACTGCTCTGAAGGCTACATTCAGAACTACAGATGCAGAGGTGATGACAGCA	852	
Dbb	61	aagtccaggaagccaggaagtcctctctctggtcgcattgcctctctctccatgtacacta	120	
QY	853	AAGTCCAGGAAGCCAGGAAGTCCTTCTCTCTGGCCATGCCCTCTTCTCATGTACACTA	912	
Dbb	121	tgctgtatttgctactatacctgcaggcccgcttcaatttgagagagccgcctgcctcc	180	
QY	913	TGCTGTATTGGTGCTATACCTGCAGGCCCGCTTCACTTGGCGAGGAGGCCGCCGTCCTC	972	
Dbb	181	ngccctcctcgtcagttcacctcttgatcatgatgagcctctcacacggagactgtctcgctat	240	
QY	973	GGCCCTCTCTCGAGTTCACCTTGATGATGATGCGCTTCTACAGGGGACTGTCTCGCGTAT	1032	
Dbb	241	cagacacacaagcacatccccagtgatgtcttgccagagatttgcctcaagagagccctgggtg	300	
QY	1033	CAGACCAACAAGCACCATCCCAAGTGATGTTCTGCACGAGATTGCTCAAGGAGCCCTGGTGG	1092	
Dbb	301	cctgctgcatagttttctt-gtgcctgacct-ttcaagactaagacga	346	
QY	1093	CCTGCTGCATAGTTTCTTCGTGCTGCACCTCTTCAAGACTAAGACGA	1140	
RESULT	14			
LOCUS	H02468	475 bp	mRNA	EST
DEFINITION	YJ35d09.r1 Homo sapiens cDNA clone 150737 5'.			20-JUN-1995
ACCESSION	H02468			
NID	9865401			
KEYWORDS	EST.			
SOURCE	human clone-150737 library=Soares placenta Mb2HP vector=PT7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13HP1 Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTCGAAGAATTCGGCGCCGAGCAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.			
ORGANISM	Homo sapiens			
REFERENCE	EuKaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chondata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 475)			
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kuchaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Ruhlberg,T., Soares,M., Tan,F., Trevasaki,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.			
TITLE	The Washu-Merck EST Project			
JOURNAL	Unpublished (1995)			
COMMENT	Contact: Wilson RK Washu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 321 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.			
FEATURES	Location/Qualifiers			
	1..475			



\*\*\*\*\*

WATERMAN

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 12:15:43 1997; MasPar time 370.56 Seconds  
Tabular output not generated. 1010.069 Million cell updates/sec

Title: >US-08-842-827-5  
Description: (1-1362) from US08842827.seq  
Perfect Score: 1362  
N.A. Sequence: 1 GCGCGAGCTGCGAAGTT.....TTTTAAAAA.....AAAAA 1362  
Comp: CCGCGTCGAGACGTTTCAA.....AAAATTTTTTTTTTTTTTT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-STS-THREE  
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13  
Database: EST-STS-FOUR  
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnSTS 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4  
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9  
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14  
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19  
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24  
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28  
103:gnEST29 104:gnEST30 105:gnEST31 106:gnSTS 107:ueEST1  
108:ueEST2

Statistics: Mean 11.582; Variance 2.791; scale 4.149

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	500	36.7	564	1	W30942	zc64f09.r1 Soares fet	0.00e+00
2	500	36.7	564	88	HS942332	zc64f09.r1 Soares fet	0.00e+00
3	382	28.0	441	15	AA192756	zq12c05.s1 Stratagene	0.00e+00
4	376	27.6	403	87	HS733324	zc64f09.s1 Soares fet	0.00e+00
5	376	27.6	403	1	W02733	zc64f09.s1 Soares fet	0.00e+00
6	336	24.7	446	12	AA181403	zp52f10.s1 Stratagene	0.00e+00
7	250	18.4	402	27	W39811	307 Mouse VM CDNA lib	0.00e+00
8	244	17.9	337	27	W39815	309 Mouse VM CDNA lib	0.00e+00
9	243	17.8	546	40	GI3351	human STS WI-12145	0.00e+00
10	225	16.5	421	8	AA126964	zl78e03.s1 Stratagene	0.00e+00
11	211	15.5	374	1	W23530	zc71h10.s1 Soares fet	4.06e-296
12	211	15.5	374	87	HS530338	zc71h10.s1 Soares fet	4.06e-296
13	190	14.0	242	27	W39815	308 Mouse VM CDNA lib	2.04e-260
14	179	13.1	236	81	HS1146241	zp52f10.r1 Stratagene	8.22e-242
15	179	13.1	236	12	AA182434	zp52f10.r1 Stratagene	8.22e-242
16	133	9.8	440	55	AA105463	mm32b06.r1 Stratagene	5.56e-165
17	124	9.1	467	10	AA016796	mh43e01.r1 Soares mou	3.42e-150
18	110	8.1	413	21	AA021853	mh86b01.r1 Soares mou	2.08e-127
19	78	5.7	385	33	AA107326	ml95b10.r1 Stratagene	9.47e-77
20	74	5.4	437	33	AA106725	mm17c02.r1 Stratagene	1.30e-70
21	46	3.4	478	26	N75714	yy31b01.r1 Soares fet	7.98e-30
22	45	3.3	914	34	AA141702	CK02248 contig Drosop	1.81e-28
23	39	2.9	422	34	AA142082	CK02216.3prime Drosop	1.47e-20
24	34	2.5	87	33	AA106892	ml85a07.r1 Stratagene	2.51e-14
25	26	1.9	569	75	AT0465	7309 Arabidopsis thal	2.53e-05
26	24	1.8	361	4	AA115689	zk36f04.s1 Soares pre	2.51e-03
27	24	1.8	533	19	AA206113	zq52g06.r1 Stratagene	2.51e-03
28	24	1.8	593	75	AT0302	4038 Arabidopsis thal	2.51e-03
29	23	1.7	73	48	HUMUT934A	Human STS UT934, 5' p	2.22e-02
30	23	1.7	157	71	AA251615	zs08h10.r1 Soares Nbh	2.22e-02
31	23	1.7	296	72	AA254840	mz75c09.r1 Soares mou	2.22e-02
32	23	1.7	296	93	MM1171194	mz75c09.r1 Soares mou	2.22e-02
33	23	1.7	345	99	MAA31464	mw38e02.r1 Soares mou	2.22e-02
34	23	1.7	345	64	AA231464	mw38e02.r1 Soares mou	2.22e-02
35	23	1.7	357	83	HS1151833	zr47e08.s1 Soares Nbh	2.22e-02
36	23	1.7	371	105	TR0691	T2504 MVAT4 bloodst	2.22e-02
37	23	1.7	403	93	MM1164661	mz58f02.r1 Soares mou	2.22e-02
38	23	1.7	431	14	AA188740	zp78d02.r1 Stratagene	2.22e-02
39	23	1.7	434	84	HS1160099	zp78d03.s1 Soares Nbh	2.22e-02
40	23	1.7	455	87	HS777332	zb2b11.s1 Soares par	2.22e-02
41	23	1.7	475	30	AA061089	ml43h09.r1 Stratagene	2.22e-02
42	23	1.7	488	6	AA121386	zn77f06.s1 Stratagene	2.22e-02
43	23	1.7	490	9	AA133331	zn92e01.s1 Stratagene	2.22e-02
44	23	1.7	519	22	AA212648	mu80c03.r1 Stratagene	2.22e-02
45	23	1.7	528	53	AA193447	zr40e04.r1 Soares Nbh	2.22e-02

ALIGNMENTS

RESULT 1  
LOCUS W30942 564 bp mRNA EST 25-NOV-1996  
DEFINITION zc64f09.r1 Soares fetal heart NBRH19W Homo sapiens CDNA clone  
327113 5' similar to WP:T28D9.3 CE02068 ;  
ACCESSION W30942  
NID Q1311934  
KEYWORDS EST.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 564)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
WashU-Merck EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 774 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 336.

Location/Qualifiers

## FEATURES

source

1..564  
/organism="Homo sapiens"  
/note="Organ: heart; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBHL19W."  
/clone="327113"  
/clone\_lib="Soares fetal heart NBHL19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
<1..>564

BASE COUNT 128 a 155 c 139 g 134 t 8 others

ORIGIN

Query Match 35.7%; Score 500; DB 1; Length 564;  
Best Local Similarity 96.6%; Pred. No. 0.00e+00;  
Matches 546; Conservative 0; Mismatches 12; Indels 7; Gaps 7;  
Db 2 agcacatcaagccttaccacccagggttttactgcaatgatgagagatcaagtaacca 61  
QY 462 AGCACCATCAAGCCTTACCACCGAGGTTTTTACTGCAATGATGAGAGCATCAAGTACCCA 521  
Db 62 ctgaaactgtgagacaataaagcgtgtcctgtgccgtggagatcgatcgcc 121  
QY 522 CTGAAACTGTGAGACAATAATGACGCTGTCTGTGCCGTGGGATCGTCAATGCC 581  
Db 122 atctctcgatcatcaccggtgggaattctaccggatctattaccctgaagaagtcgggtcc 181  
QY 582 ATCTCTCGGATCATCAGCGGGAATTCTACCGGATCTATTACCTGAAGAAGTCGGGTC- 640  
Db 182 gacgattcagaacccctacgtggcagcactctataagcaagtggactctctctcttgg 241  
QY 641 GACGATTGAGAACCCCTACGTGGGAGGACTCTATAGCAAGTGGGCTGCTCTCTTTGG 700  
Db 242 cctgtgc-atcagccagtgctttcacagacattgccaagtgtccatagggcgctcgctcc 300  
QY 701 CTGTGCCATCAGCCAGTCTTTCACAGACATTTGCCAAGTGTCCATAGGCGGCTCGCTCC 760  
Db 301 tcactcttgagtgctgcaacctgatttcagccagatcaactgctcgaagctaca 360  
QY 761 TCACCTCTGA-GTGTCTGCACCCCTGATTTACGCCAGATCAACTGCTCTGAGGCTACA 819  
Db 361 ttccagaatcacagatgcagaggtgatgacgcaagtcacaggangccaggaagtccttcn 420  
QY 820 TTCCAGAACTACAGATGCAGAGGTGATGACAGCAAAAGTCCAGGAAGCCAGGAAGTCTCTCT 879  
Db 421 tctctggccatgcctctctctccatgtacactatgctgggaatttgctatancctcag 480  
QY 880 TCTCTGGCCATGCTCTCTCTCCATGTACACTATGTGTA-FTTGGTGTATACCTGAG 938  
Db 481 gcccgcttcaattggcagagacc-gnctgctnccggccctcctcgagctnacct-gatn 538  
QY 939 GCCCGCTTCACTTGGCGAGGAGCCCGCTCTCCGCCCTCTCTCGGCTTACCTGAGTTCACCTTGAT 998

Db 539 atgatggcctntacaacgggactg 563  
QY 999 ATGATGGCCTTCTACA-CGGGACTG 1022

## RESULT 2

ID HS942332 standard; RNA; EST; 564 BP.  
AC W30942;  
NI Q1311934  
DT 13-MAY-1996 (Rel. 47, Created)  
DE 07-MAR-1997 (Rel. 51, Last updated, Version 2)  
DE zc64f09.r1 Soares fetal heart NBHL19W Homo sapiens cDNA clone  
DE 327113 5' similar to WP:T28D9.3 CE02068 ;  
KW EST.  
OS Homo sapiens (human)  
OC Eukaryotae; mitochondria; eukaryotes; Metazoa; Chordata;  
OC Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP 1-564  
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
RA Trevasakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;  
RT "The WashU-Merck EST Project";  
RL Unpublished  
CC Contact: Wilson RK WashU-Merck EST Project Washington University  
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:  
CC estewatson.wustl.edu This clone is available royalty-free through  
CC LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for  
CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:  
CC mob.REGA+ET High quality sequence stop: 336.  
FH Key Location/Qualifiers  
FT source 1..564  
FT /organism="Homo sapiens"  
FT /note="Organ: heart; Vector: pT73D (Pharmacia) with a  
FT modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
FT strand cDNA was primed with a Not I - oligo(dT) primer [5'  
FT TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTTTTTTTTT 3'],  
FT double-stranded cDNA was size selected, ligated to Eco RI  
FT adapters (Pharmacia), digested with Not I and cloned into  
FT the Not I and Eco RI sites of a modified pT73 vector  
FT (Pharmacia). Library went through one round of  
FT normalization to a Cot = 5. Library constructed by M.Fatima  
FT Bonaldo. This library was constructed from the same fetus  
FT as the fetal lung library, Soares fetal lung NBHL19W."  
FT /clone="327113"  
FT /clone\_lib="Soares fetal heart NBHL19W"  
FT /sex="unknown"  
FT /dev\_stage="19 weeks"  
FT /lab\_host="DH10B (ampicillin resistant)"  
FT <1..>564  
SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;

Query Match 35.7%; Score 500; DB 88; Length 564;

Best Local Similarity 96.6%; Pred. No. 0.00e+00;

Matches 546; Conservative 0; Mismatches 12; Indels 7; Gaps 7;

Db 2 agcacatcaagccttaccacccagggttttactgcaatgatgagagatcaagtaacca 61

QY 462 AGCACCATCAAGCCTTACCACCGAGGTTTTTACTGCAATGATGAGAGCATCAAGTACCCA 521

Db 62 ctgaaactgtgagacaataaagcgtgtcctgtgccgtggagatcgatcgcc 121

QY 522 CTGAAACTGTGAGACAATAATGACGCTGTCTGTGCCGTGGGATCGTCAATGCC 581

Db 122 atctctcgatcatcaccggtgggaattctaccggatctattaccctgaagaagtcgggtcc 181

QY 582 ATCTCTCGGATCATCAGCGGGAATTCTACCGGATCTATTACCTGAAGAAGTCGGGTC- 640



FT adapters (Pharmacia), digested with Not I and cloned into  
 FT the Not I and Eco RI sites of a modified pT73 vector  
 FT (Pharmacia). Library went through one round of  
 FT normalization to a Cot = 5. Library constructed by M. Fatim

a  
 FT as the fetal lung library. Soares fetal lung NBHL19W."  
 FT /clone="327113"  
 FT /clone\_lib="Soares fetal heart NBHL19W"  
 FT /sex="unknown"  
 FT /dev\_stage="19 weeks"  
 FT /lab\_host="DH10B (ampicillin resistant)"  
 FT complement(<1.>403)  
 FT SQ Sequence 403 BP; 106 A; 87 C; 118 G; 91 T; 1 other;

Query Match 27.6%; Score 376; DB 87; Length 403;  
 Best Local Similarity 99.0%; Pred. No. 0.00e+00;  
 Matches 394; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Db 1 taaaatcagtcgggcaaaagtgtttccctacattctctgtctgtgatgagattggagagc 60  
 Cp 1347 TAAATAATCAGTCGGGCAAAAGTGTTCCTACATTCCTACTGCTGATGAGATTGGAGAGC 1288

Db 61 agcaagaacttctgtcagcagtcattttcaaaaacagctcagggagtggtgcacct 120  
 Cp 1287 AGCAGAAGTTCGTCAGCAGTCATTTTACAAAACAGCTCAGGAGGTGGTGCACTT 1228

Db 121 acatcatgttgggtgattgtctctgctcaaatatgtccacaggtgaaaggattctctcc 180  
 Cp 1227 ACATCATGTTGGTGTGATTGTTCCTGTCAATAATGTCCACAGGTGAAGGATTCCTCTCC 1168

Db 181 ggtatgacaggggacggcagagcgctgccttagcttgaaagagtcagacagaaga 240  
 Cp 1167 GGATACAGGGGACGAGGAGAGCGTCTCTTAGCTCTTGAGAGGTCAGACAGAAGA 1108

Db 241 aaactatgcagcagggccaccgggctccttgagcaaatcctgcagaacatcactgggat 300  
 Cp 1107 AAATATGCAGAGGCCACCGAGGCTCCTTGAGCAAACTCTGCAGAACTCCTCAGATCAGTGGAT 1048

Db 301 ggtgcttggtgctgattacgcgagacagtcctcgctgtagaagccatcatgcaagtg 360  
 Cp 1047 GGTGCTTGTTGGTCTGAT -ACCGAGACAGTCCCGGTGAGAGCCATCATGATCAGGTG 989

Db 361 aactgcagggggccggagcaggcggggctctctgcg 398  
 Cp 988 AACT-GCAGGAGGGCCGGAGCGGGG-CTCCTCGC 953

RESULT 5  
 LOCUS w02733 403 bp mRNA EST 25-NOV-1996  
 DEFINITION z64f09.s1 Soares fetal heart NBHL19W Homo sapiens cDNA clone  
 327113 3'  
 ACCESSION w02733  
 NID 91274792  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 403)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevas, E., Waterston, R., Williamson, A., Wohldmann, P. and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800

Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 774 Std Error: 0.00  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 312.  
 Location/Qualifiers  
 1..403  
 /organism="Homo sapiens"  
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGCGCGCATCTTTTTTTTTTTTTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M. Fatima Bonaldo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NBHL19W."  
 /clone="327113"  
 /clone\_lib="Soares fetal heart NBHL19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 complement(<1.>403)

## FEATURES

source  
 mRNA  
 BASE COUNT 106 a 87 c 118 g 91 t 1 others  
 ORIGIN  
 Query Match 27.6%; Score 376; DB 1; Length 403;  
 Best Local Similarity 99.0%; Pred. No. 0.00e+00;  
 Matches 394; Conservative 0; Mismatches 1; Indels 3; Gaps 3;  
 Db 1 taaaatcagtcgggcaaaagtgtttccctacattctctgtctgtgatgagattggagagc 60  
 Cp 1347 TAAATAATCAGTCGGGCAAAAGTGTTCCTACATTCCTACTGCTGATGAGATTGGAGAGC 1288  
 Db 61 agcaagaacttctgtcagcagtcattttcaaaaacagctcagggagtggtgcacct 120  
 Cp 1287 AGCAGAAGTTCGTCAGCAGTCATTTTACAAAACAGCTCAGGAGGTGGTGCACTT 1228  
 Db 121 acatcatgttgggtgattgtctctgctcaaatatgtccacaggtgaaaggattctctcc 180  
 Cp 1227 ACATCATGTTGGTGTGATTGTTCCTGTCAATAATGTCCACAGGTGAAGGATTCCTCTCC 1168  
 Db 181 ggtatgacaggggacggcagagcgctgccttagcttgaaagagtcagacagaaga 240  
 Cp 1167 GGATACAGGGGACGAGGAGAGCGTCTCTTAGCTCTTGAGAGGTCAGACAGAAGA 1108  
 Db 241 aaactatgcagcagggccaccgggctccttgagcaaatcctgcagaacatcactgggat 300  
 Cp 1107 AAATATGCAGAGGCCACCGAGGCTCCTTGAGCAAACTCTGCAGAACTCCTCAGATCAGTGGAT 1048  
 Db 301 ggtgcttggtgctgattacgcgagacagtcctcgctgtagaagccatcatgcaagtg 360  
 Cp 1047 GGTGCTTGTTGGTCTGAT -ACCGAGACAGTCCCGGTGAGAGCCATCATGATCAGGTG 989  
 Db 361 aactgcagggggccggagcaggcggggctctctgcg 398  
 Cp 988 AACT-GCAGGAGGGCCGGAGCGGGG-CTCCTCGC 953  
 RESULT 6  
 LOCUS AA181403 446 bp mRNA EST 06-JAN-1997  
 DEFINITION zp52ff10.s1 Stragatene HeLa cell s3 937216 Homo sapiens cDNA clone  
 613099 3'  
 ACCESSION AA181403  
 NID 91765128  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE AUTHORS	TITLE JOURNAL COMMENT	CONTACT	FEATURES	BASE SEQUENCE ORIGIN	Query Match Best Local Similarity Matches	Score 99.4%; 342;	DB No. 0.00e+00; 0;	Length 446; Mismatches	Indels 1;	Gaps 1;
Eukaryotas; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarhini; Homiidae; Homo. 1 (bases 1 to 446)		WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu	This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 286. Location/Qualifiers 1..446 /organism="Homo sapiens" /note="vector: paluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XPR vector. -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"/clone.lib="613099" /sex="female" /dev_stage="HeLa S3 cell line" /lab_host="SOLR (kanamycin resistant)" complement(<1..>446)	111 a 103 c 125 g 101 t 6 others	24.7%; 99.4%; 342;	DB No. 0.00e+00; 0;	Length 446; Mismatches 1;	Indels 1;	Gaps 1;	
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu										
WashU-Merck EST Project Unpublished (1995)										
WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314										

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus.
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS	Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
TITLE	Mus.
JOURNAL	1 (bases 1 to 402)
COMMENT	Stewart,G.J., Savioz,A. and Davies,R.W. Sequence analysis of 497 mouse brain ESTs expressed in the substantia nigra Genomics 39, 147-153 (1997)  Contact: Davies,R.W. Robertson Laboratory of Biotechnology Institute of Biomedical and Life Sciences, Division of Molecular Genetics Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU Tel: 44 141 330 5102 Fax: 44 141 330 5102/4878 Email: gbg21@udcf.gla.ac.uk Seq primer: T7/T3alpha High quality sequence stop: 422. Location/Qualifiers 1..402 /oranism="Mus musculus" /note="Organ: Brain; Vector: pSPORT1; Mouse Ventral Midbrain directional cDNA library in pSPORT1. The library was created by subtractive hybridisation using VM mRNA as the target nucleic acid population and total cerebellar ss cDNA as the driver nucleic acid population. The clones are 3' directed." /clones="smp22 1.22" /clone_lib="Mouse VM cDNA library" /tissue_type="ventral midbrain" <1..>402
BASE COUNT	69 a 131 c 101 g 101 t
ORIGIN	
Query Match	18.4%; Score 250; DB 27; Length 402;
Best Local Similarity	84.4%; Pred. No. 0.00e+00;
Matches 313; Conservative	0; Mismatches 57; Indels 1; Gaps 1;
Db	2 ggccagcgctcttcttcactggttcactatgctgtatctgtgtgtctacattcaggccgcg 61
QY	885 GGCCATGCTCCTCTTCCCATGTACACTATGTGTATTGGTGTTATACCTGCAGGCCCGC 944
Db	62 ttcaccttggcggggccgacctgctecgccctctctcagtctcatctttgctcatgatg 121
QY	945 TTCACITGGGAGGAGCCGCGCTGCTCGGGCCCCCTCTCGAGTTCACCTTGATCATGATG 1004
Db	122 gcctttcaacgggattgtcacgggtatctgaactacaagaatcatcctcagcgatgctctg 181
QY	1005 GCCTTCTACAGGGGACTGTCTGGCGTATCAGACCACAAGCACCATCCAGTAGTGTCTG 1064
Db	182 gcaggatttgccaaaggagctctgtggcctgctgcattgcttcttcgttcagaccctc 241
QY	1065 GCAGGATTGCTCAAGGAGCCCTGGTGCCCTGCTGCATAGTTTTTCITCGTGTCTGACCTC 1124
Db	242 ttcaagactaagaagagcctctcactggcgccccctcgatcagggaggagatccttgtct 301
QY	1125 TTCAAAGCTAAGACGACGCTCTCCCTGGCTGCCCTGCTATCGCGGAAGGAAATTCCTTTCA 1184
Db	302 ccgctggacatcatcgccagggaacaatcacatcaccattgcttagatgctcgggcctcc 361
QY	1185 CCTGTGGACATATTGACAGG-AACAATCACCACAACATGATGTAGTGGCCACCCACCTC 1243
Db	362 cggagcgcttt 372
QY	1244 CTGAGCTGTGTT 1254
RESULT LOCUS	8 W39816 337 bp MRNA EST 05-FEB-1997

```

DEFINITION 309 Mouse VM cDNA library Mus musculus cDNA clone smp22 1.25.
ACCESSION W39816
NID G1816979
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
JOURNAL substantia nigra
COMMENT Genomics 39, 147-153 (1997)
Contact: Davies,R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gba21eudcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence stop: 337.
Location/Qualifiers
FEATURES
source
1..337
/organism="Mus musculus"
/note="Organ: brain; Vector: pSPORT1; Mouse Ventral
Midbrain directional cDNA library in pSPORT1. The library
was created by subtractive hybridisation using VM mRNA as
the target nucleic acid population and total cerebellar ss
cDNA as the driver nucleic acid population. The clones are
3' directed."
/clone="smp22 1.25"
/clone_lib="Mouse VM cDNA library"
/tissue_type="ventral midbrain"
BASE COUNT 59 a 112 c 83 g 83 t
ORIGIN
Query Match 17.9%; Score 244; DB 27; Length 337;
Best Local Similarity 87.4%; Pred. No. 0.00e+00;
Matches 291; Conservative 0; Mismatches 41; Indels 1; Gaps 1;
Db 3 ggccagcctcttccatgttccactatgctgtatctgtctacattcagggccgc 62
QY 885 ggccatgcctcttccatgttccactatgctgtatctgtctacattcagggccgc 944
Db 63 ttccactgcgcggggccgactctgcgcgcctcctgcagttcactttgctcatgatg 122
QY 945 ttccactgcgcggggccgactctgcgcgcctcctgcagttcactttgctcatgatg 1004
Db 123 gcccttacacggggtgttcacgggtatctgactacacgacatcctcagcgtctcgtg 182
QY 1005 gcccttacacggggtgttcacgggtatctgactacacgacatcctcagcgtctcgtg 1064
Db 183 gcagattggccacggagccttggtgcctgcctgcctgcctgcctgcctgcctgcctgc 242
QY 1065 gcagattggccacggagccttggtgcctgcctgcctgcctgcctgcctgcctgcctgc 1124
Db 243 ttcaagactaaagagccttccactgcctgcctgcctgcctgcctgcctgcctgcctgcct 302
QY 1125 ttcaagactaaagagccttccactgcctgcctgcctgcctgcctgcctgcctgcctgcct 1184
Db 303 cccgtgacatcatgccagggaacaatcacca 335
QY 1185 cctgtggacattattgacagg-acacatcacca 1216
RESULT 9
LOCUS G13351
DEFINITION human STS WI-12145.

```

```

ACCESSION G13351
NID G1277460
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene
collection.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Cathartini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
JOURNAL Mapped STS
COMMENT Unpublished (1995)
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: AAATCAGTCGGGCAAAAGT
Primer B: TATTGACAGGAACAATCACCAC
STS size: 150
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3
FEATURES
Derived from dbEST (genbank accession T71976).
source
location/Qualifiers
STS
/organism="Homo sapiens"
primer_bind 1..546
/map="981.7 cR from top of Chr2 linkage group"
primer_bind 1..150
/map="981.7 cR from top of Chr2 linkage group"
complement(129..150)
primer_bind 148 a 109 c 147 g 128 t 14 others
BASE COUNT
ORIGIN
Query Match 17.8%; Score 243; DB 40; Length 546;
Best Local Similarity 98.4%; Pred. No. 0.00e+00;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 1 aaaaacatcgccgcaaaagtcttcctacattctcactctcactcagatggagagcag 60
Cp 1345 AAATCAGTCGGGCAAAAGTTCCTTCTACTGCTGATGAGATTGGAGCGAG 1286
Db 61 caagaacttgctgcagcagtcattttacaaaaaacagctcaggagtggtggccacctac 120
Cp 1285 CAAGAACTTGCTGTCAGCAGTCATTTTACAAAAAACAGCTCAGGAGTGGTGACCTAC 1226
Db 121 atcaatgtgtggtgattgttcctcgtcgaataatgtccacaggtgaaaggatttccttcgg 180

```







```

RESULT 14
ID HS1146241 standard; RNA; EST; 236 BP.
AC AA182434;
NI 91766132
DT 24-FEB-1997 (Rel. 51, Created)
DE 2652f10.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
DE 613099 5' similar to TR:G1161100 G1161100 HYDROGEN
DE PEROXIDE-INDUCIBLE PROTEIN ;
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP 1-236
RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
RA Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,
RA Waterston R., Williamson A., Wohlmann P., Wilson R.;
RT "WashU-Merck EST Project";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC estevaton.wustl.edu This clone is available royalty-free through
CC LNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Seq primer: -28M13 rev2 from Amersham High
CC quality sequence stop: 91.
FH Key Location/Qualifiers
FT 1..236
FT /organism="Homo sapiens"
FT /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
FT XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3
FT epithelioid carcinoma cells grown to semi-confluency
FT without induction. Average insert size: 1.5 kb; Uni-ZAP XR
FT Vector. -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3'
FT adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'
FT /clone="613099"
FT /sex="female"
FT /dev_stage="HeLa S3 cell line"
FT /lab_host="SOLR (kanamycin resistant)"
FT <1..>236
FT mRNA
FT BASE COUNT 60 a 59 c 52 g 56 t 9 others
FT ORIGIN
FT
FT Query Match 13.1%; Score 179; DB 81; Length 236;
FT Best Local Similarity 91.1%; Pred. No. 8.22e-242;
FT Matches 216; Conservative 0; Mismatches 16; Indels 5; Gaps 5;
FT
FT 1 ttccagaccctacgtggcagcactctataagcaagtggctgctnctnttttggctgt- 59
FT QY 646 TTCAGAACCCCTACGTGGCAGACCTCTATAGCAAGTGGGCTGCTCTCTTTGGCTGTG 705
FT
FT 60 catcaagccagcttttcacagacattgccaaagtgtccataggcgctgcgtcctcaact 119
FT Db
FT QY 706 CCATCAGCCAGCTTTTCACAGACATTGCCAAAGTGTCCATAGGCGGCTCGCTCCTCACT 765
FT
FT 120 tcttgagtgtctnaacccctgatttnaagccagattgaactgtcctgaaggctacatnca 179
FT QY
FT 766 TCTTGAGTGTCTGCACACCTGATTTCACGCCAGATCAACT-GCTCTGAAGGCTACATTCA 823
FT Db
FT 180 gaacttacagatgnagaggtgatgacagcaaatgccaggaagccaaaggaagtccttc 236
FT Db
FT QY 824 GAACACT-ACAGATGCAGAGGTGATGACAGCAAAAGTCCAGGAAGCCA-GGAAGTCTCTTC 878
FT
FT Search completed: Thu Nov 6 12:23:06 1997
FT Job time : 443 secs.

```

```

91766132
EST.
human.
Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 236)
Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,
Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,
Parsons J., Rifkin L., Rohlfing T., Tan F., Trevaskis E.,
Waterston R., Williamson A., Wohlmann P. and Wilson R.
WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevaton.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 91.
Location/Qualifiers
1..236
/organism="Homo sapiens"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector. -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3'
adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'
/clone="613099"
/sex="female"
/dev_stage="HeLa S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
<1..>236
mRNA
BASE COUNT 60 a 59 c 52 g 56 t 9 others
ORIGIN
Query Match 13.1%; Score 179; DB 12; Length 236;
Best Local Similarity 91.1%; Pred. No. 8.22e-242;
Matches 216; Conservative 0; Mismatches 16; Indels 5; Gaps 5;
1 ttccagaccctacgtggcagcactctataagcaagtggctgctnctnttttggctgt- 59
QY 646 TTCAGAACCCCTACGTGGCAGACCTCTATAGCAAGTGGGCTGCTCTCTTTGGCTGTG 705
Db 60 catcaagccagcttttcacagacattgccaaagtgtccataggcgctgcgtcctcaact 119
QY 706 CCATCAGCCAGCTTTTCACAGACATTGCCAAAGTGTCCATAGGCGGCTCGCTCCTCACT 765
Db 120 tcttgagtgtctnaacccctgatttnaagccagattgaactgtcctgaaggctacatnca 179
QY 766 TCTTGAGTGTCTGCACACCTGATTTCACGCCAGATCAACT-GCTCTGAAGGCTACATTCA 823
Db 180 gaacttacagatgnagaggtgatgacagcaaatgccaggaagccaaaggaagtccttc 236
QY 824 GAACACT-ACAGATGCAGAGGTGATGACAGCAAAAGTCCAGGAAGCCA-GGAAGTCTCTTC 878
Search completed: Thu Nov 6 12:23:06 1997
Job time : 443 secs.

```

```

RESULT 15
LOCUS AA182434 236 bp mRNA EST 07-JAN-1997
DEFINITION zp52f10.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
613099 5' similar to TR:G1161100 G1161100 HYDROGEN
PEROXIDE-INDUCIBLE PROTEIN ;
ACCESSION AA182434

```

**This Page Blank (uspto)**

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:27:48 1997; MasPar time 7.48 Seconds  
Tabular output not generated.  
511.387 Million cell updates/sec  
Title: >US-08-842-827-6  
Description: (1-311) from US08842827.pap  
Perfect Score: 2376  
Sequence: 1 MQNKKDKAIVPSKNGGSP.....RKEILSPVDIDRNNHHNM 311  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

\*\*\*\*\*  
MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:27:48 1997; MasPar time 7.48 Seconds  
Tabular output not generated.  
511.387 Million cell updates/sec  
Title: >US-08-842-827-6  
Description: (1-311) from US08842827.pap  
Perfect Score: 2376  
Sequence: 1 MQNKKDKAIVPSKNGGSP.....RKEILSPVDIDRNNHHNM 311  
Scoring table: PAM 150  
Gap 11  
Searched: 101610 seqs, 12294212 residues  
Post-processing: Minimum Match 08  
Listing first 45 summaries  
Database: a-geneseq28  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21  
Statistics: Mean 33.923; Variance 140.094; scale 0.242  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	4.8	352	20	W01663	7.64e-01
2	111	4.8	352	11	R58712	7.64e-01
3	110	4.7	251	21	W20277	9.15e-01
4	96	4.1	111	11	R57067	1.07e-01
5	95	4.1	492	2	R11360	1.27e-01
6	95	4.1	836	17	R87153	1.27e-01
7	95	4.1	836	11	R58912	1.27e-01
8	95	4.1	904	17	R87147	1.27e-01
9	95	4.1	904	11	R58907	1.27e-01
10	93	4.0	370	11	R57066	1.27e-01
11	90	3.9	583	20	W06891	2.94e-01
12	90	3.9	583	18	R97230	2.94e-01
13	89	3.8	263	4	R25424	3.47e-01
14	89	3.8	370	12	P58686	3.47e-01
15	88	3.8	588	19	W03991	4.10e-01
16	86	3.7	88	16	R90991	5.69e-01
17	87	3.7	121	21	W20596	4.83e-01
18	87	3.7	121	15	R75902	4.83e-01
19	87	3.7	357	9	R45847	4.83e-01
20	86	3.7	357	12	R58685	5.69e-01

21	85	3.7	357	9	R45848	Human 5HT5a serotonin	6.70e+01
22	87	3.7	377	21	W20561	H. pylori transmembra	4.83e+01
23	85	3.7	829	16	R86578	Autotaxin derived fro	6.70e+01
24	85	3.7	849	16	R86595	N-tera 2D1 autotaxin	6.70e+01
25	85	3.7	915	16	R86596	A2058 autotaxin prote	6.70e+01
26	84	3.6	329	13	R74037	MHC polypeptide HLA D	7.87e+01
27	83	3.6	324	20	W11860	Human proteosome subu	9.25e+01
28	83	3.6	369	7	R39261	Human somatostatin re	9.25e+01
29	83	3.6	369	7	R39262	Murine somatostatin r	9.25e+01
30	83	3.6	369	18	R97269	Human somatostatin re	9.25e+01
31	83	3.6	455	3	R12362	Octopus rhodopsin mem	9.25e+01
32	83	3.6	570	18	R97231	Stem cell marker HCAP	9.25e+01
33	84	3.6	724	5	R27648	Human calcium channel	9.25e+01
34	84	3.6	1138	2	R06461	Human calcium channel	9.25e+01
35	84	3.6	1138	9	R46226	BtPES1245 protoxin.	7.87e+01
36	84	3.6	1138	7	R37214	Bacillus thuringiens	7.87e+01
37	84	3.6	1138	9	R46225	B.t. toxin HD867	7.87e+01
38	84	3.6	1138	7	R37213	Bacillus thuringiens	7.87e+01
39	83	3.6	2251	14	R71009	B.t. toxin HD511	7.87e+01
40	83	3.6	2270	14	R71010	Human neuronal calciu	9.25e+01
41	83	3.6	2270	13	R69604	Human neuronal calciu	9.25e+01
42	83	3.6	2458	2	R07640	Calcium channel alpha	9.25e+01
43	83	3.6	2458	2	R04031	Deduced protein sequ	9.25e+01
44	83	3.6	4472	19	R97245	Full length T4 encode	9.25e+01
45	82	3.5	368	1	R06606	Virulence gene cluste	9.25e+01
						Lipase encoded by rec	1.09e+02

ALIGNMENTS

RESULT 1  
ID W01663 standard; Protein; 352 AA.  
AC W01663;  
DE 01-APR-1997 (first entry)  
DE P(H218); G-protein coupled receptor.  
KW P(H218); G-protein coupled receptor; cell differentiation; proliferation;  
KW proline directed kinase; cell division; growth factor response; rat-edg;  
KW therapy; diagnosis.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT Modified site 19  
FT /note- "predicted N-glycosylation site"  
FT Binding\_site 222..225  
FT /note- "possible nucleotide binding site"  
FT Binding\_site 309..315  
FT /note- "possible nucleotide binding site".  
FT Misc\_difference 342..343  
FT /note- "consensus sequence for proline directed  
FT kinases"  
PN US5585476-A.  
PD 17-DEC-1996.  
PF 15-FEB-1994; 196989.  
PR 15-FEB-1994; US-196989.  
PA (MACL/) MACLENNAN A J.  
PI Macleennan AJ;  
DR WPI; 97-051235/05.  
DR N-PSDB; T58505.  
PT DNA encoding rat protein p(H218) - associated with cell  
PT proliferation and/or differentiation  
PS Claim 1; Column 17-20; 33pp; English.  
CC This sequence is a novel rat protein p(H218), a member of the G-protein  
CC coupled receptor (GPR) superfamily of proteins. Several features of  
CC p(H218) are common to all other GPRs, including: (i) seven regions of  
CC hydrophobicity which are predicted to act as membrane spanning domains;  
CC (ii) a consensus sequence for N-linked glycosylation in its predicted  
CC N-terminal extracellular domain, and (iii) a conserved cysteine residue  
CC and several serine and threonine residues in its predicted C-terminal  
CC domain. In addition, p(H218) contains many other residues which are  
CC highly conserved among most GPRs. However, p(H218) is distinct from these  
CC GPRs in that it does not contain certain highly conserved residues.  
CC Perhaps most notable are the aspartate and tyrosine residues at the  
CC cytoplasmic end of the third transmembrane domain, and the cysteine  
CC residue at the extracellular end of the same transmembrane domain. The  
CC amino acid similarity between p(H218) and p(rat-edg) (W01664) suggests



```

Query Match          4.1%; Score 96; DB 11; Length 111;
Best Local Similarity 32.7%; Pred. No. 1.07e+01;
Matches 16; Conservative 15; Mismatches 16; Indels 2; Gaps 2;

Db 31 eqkerramvlllgvfwlcpfflttlspicacslp-p-ikwsif 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 248 DKKHPSDVLGFAOGALVACCVFVSDLFKTKTTLSPAPAIRKEIL 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID R11360 standard; Protein; 492 AA.
AC R11360;
DT 04-JUN-1991 (first entry)
DE Glucose Transporter Protein from CHO cells.
KW GTP; serum-independence; Chinese Hamster Ovary;
OS Insulin-independence.
KW Cricetulus sp.
PN WQ9103554-A.
PD 21-MAR-1991.
PF 28-AUG-1990; U04849.
PR 01-SEP-1989; US-402204.
PR 20-JUN-1990; US-541426.
PA (GETH ) GENENTECH INC.
PI Thomas JN, Williams SR;
PI WPI: 91-102073/14.
DR N-PSDB; Q11148.
PT Transfecting eukaryotic cells with a replicable vector -
PT comprising nucleic acid sequence encoding a GTP operably linked
PT to control sequences recognised by transfected cells
PS Claim 27; Fig 11; 58pp; English.
CC A cDNA library was prepared in lambda gt10 from total cellular RNA
CC from CHO-FR4 cells. The library was screened with 3 probes (see
CC Q11149-Q11151) based on highly conserved regions of the known rat
CC and human facilitative GTP sequences. One positive plaque was
CC identified (out of 150000), replated and screened with 2 of the 3
CC probes. Phage DNA was prepared from one of the positive plaques and
CC its 2.5kb cDNA insert excised using EcoRI. The insert was ligated
CC into pRK5 and transformed into E.coli strain SR101. Recombinant
CC clones were picked for sequencing. The complete GTP-encoding
CC sequence was used to transform eukaryotic cells to allow them to be
CC grown in culture medium free of expensive protein growth factors,
CC fatty acids, metal ions, transferrin, serum and insulin.
CC Sequence 492 AA;

Query Match          4.1%; Score 95; DB 2; Length 492;
Best Local Similarity 28.6%; Pred. No. 1.27e+01;
Matches 20; Conservative 18; Mismatches 26; Indels 6; Gaps 5;

Db 44 yntwtvhrvgeipvtlttlswlsvaifvgmigsfvglfv-rfgrnslmnmll 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 YIQNYRCRGGDSKVGQARKSFPS-GHASFMTML--YLV-LYLQARFTWRGALLRPLL 230
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 afvsvlmgfv 112
      : : : : :
QY 231 QF-TLIWMAF 239

RESULT 6
ID R87153 standard; Protein; 836 AA.
AC R87153;
DT 29-AUG-1996 (first entry)
DE Alternatively spliced pc43 #1.
KW Protocadherin; pc3; pc4; human; rat; cadherin; cell adhesion; mouse;
KW catenin; therapy; clone; frog; fruit fly.
OS Homo sapiens.
PN WQ9600289-AL.
PD 04-JAN-1996.
PF 26-JUN-1995; U08071.
PR 27-JUN-1994; US-268161.
PA (DOHE-) DOHENY EYE INST.
PI Suzuki S;
PI WPI: 96-068873/07.

```

---

```

DR N-PSDB; T03624.
PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT pc5 - involved in cell-cell adhesion and regulation activities
PS Example 3; Page 100-104; 146pp; English.
CC R87153 and R87154 represent possible alternative
CC the clone of the human protocadherin sequence, designated pc43. The cDNA
CC encoding these sequences were isolated after screening a human foetal
CC brain cDNA library (contained within lambdaZAPIII vectors), with 32P
CC labelled versions of the sequences represented by T03605 and T03606. The
CC most abundant spliced version of pc43 is represented in T03622. The
CC cytoplasmic domain of cadherin interacts with the cytoskeleton through
CC catenins and other cytoskeleton associated proteins. The cytoplasmic
CC domain is not present in all cadherins, but in those which possess it, it
CC is essential for the cadherins adhesive function. The cadherins which do
CC not possess a cytoplasmic domain appear to function via a different
CC method from those with a cytoplasmic domain. These protein sequences are
CC involved in cell-cell adhesion. This sequence may have regulatory
CC functions in the cell, as well as the cell-cell adhesive properties.
CC Antibodies produced against these sequences are useful for modulating the
CC binding activity of protocadherins, and can be used therapeutically.
SQ Sequence 836 AA;

Query Match          4.1%; Score 95; DB 17; Length 836;
Best Local Similarity 33.3%; Pred. No. 1.27e+01;
Matches 15; Conservative 15; Mismatches 11; Indels 4; Gaps 3;

Db 698 ilvsvgfvtvfgvlii--f-kvykwkgsrdlyrap-vsslyrtppg 738
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 VLCAGVITAILAITGTEFYIYIKRSRSTIQNPYVALYKQVG 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID R58912 standard; Protein; 836 AA.
AC R58912;
DT 17-APR-1995 (first entry)
DE Product of alternative splicing of human protocadherin-43 mRNA.
KW Cadherin; protocadherin; cell adhesion molecule.
OS Homo sapiens.
PN WQ9414960-A.
PD 07-JUL-1994.
PF 23-DEC-1993; U12588.
PR 29-DEC-1992; US-998003.
PA (DOHE-) DOHENY EYE INST.
PI Suzuki S;
PI WPI: 94-293849/36.
DR N-PSDB; Q69000.
PT Polynucleotide sequences encoding new proto:cadherins - useful
PT for modulating natural binding and regulating activities.
PS Example; Page 96-99; 114pp; English.
CC Two full length human cDNAs corresp. to the partial cDNAs of
CC HUMAN-42 and HUMAN-43 (Q68981,Q68982) were isolated
CC from human fetal brain cDNA library. Several overlapping cDNA
CC clones were isolated with each probe including two cDNAs which
CC contained the putative entire coding sequences of two novel
CC proteins designated protocadherin-42 (pc42) and protocadherin-43
CC (pc43). The DNA and deduced AA sequences of pc42 are in
CC Q68997/R58906, while those of pc43 are in Q68998/R58907. Sequence
CC analysis of various overlapping protocadherin cDNA clones revealed
CC that some clones contd. unique sequences at the 3' end. The
CC sequences forming the boundaries of the 3' end regions are
CC consistent with the consensus sequence of mRNA splicing, suggesting
CC that these clones may corresp. to alternatively spliced mRNAs.
CC The DNA and AA sequences of one possible product of alternative
CC splicing of pc42 mRNA are given in Q68999/R58911. The DNA and AA
CC sequences of two possible products of alternative splicing of pc43
CC mRNA are respectively presented in Q69000/R58912 and Q69001/R49144.
SQ Sequence 836 AA;

Query Match          4.1%; Score 95; DB 11; Length 836;
Best Local Similarity 33.3%; Pred. No. 1.27e+01;
Matches 15; Conservative 15; Mismatches 11; Indels 4; Gaps 3;

Db 698 ilvsvgfvtvfgvlii--f-kvykwkgsrdlyrap-vsslyrtppg 738

```











**This Page Blank (uspto)**

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	187	8.0	274	11	S69561	2.09e-12
2	167	7.2	289	11	S70114	1.97e-09
3	138	5.9	216	10	S32217	2.46e-05
4	128	5.5	157	14	S28067	5.33e-04
5	122	5.2	185	14	S58936	3.20e-03
6	120	5.2	1571	11	S50659	5.75e-03
7	118	5.1	239	12	S64377	1.03e-02
8	111	4.8	180	14	E48909	7.57e-02
9	111	4.8	352	14	E31465	7.57e-02
10	112	4.8	407	5	S26033	5.72e-02
11	109	4.7	159	13	S37995	1.32e-01
12	99	4.3	236	16	I51740	1.94e+00
13	99	4.3	236	16	I51745	1.94e+00
14	99	4.3	236	16	I51741	1.94e+00
15	101	4.3	346	12	S49959	1.15e+00
16	101	4.3	381	8	S24611	1.15e+00
17	101	4.3	404	8	S21660	1.15e+00
18	101	4.3	404	4	LABECA	1.15e+00
19	101	4.3	529	13	S12787	1.15e+00
20	99	4.3	595	10	E64460	1.94e+00
21	101	4.3	1612	11	S51243	1.15e+00

ORGANISM	#formal_name	Saccharomyces cerevisiae
DATE	24-Aug-1996	#sequence_revision 06-Sep-1996 #text_change 06-Sep-1996

```

ACCESSIONS      S70114
REFERENCE        S70114
#authors        Fulton, L.
#submission     submitted to the EMBL Data Library, May 1996
#description     The sequence of S. cerevisiae cosmid 9819.
#accession      S70114
#molecule_type DNA
#residues       1-289 #label FULL
#cross-references EMBL:U51031

GENETICS

#map_position 4R
SUMMARY         #length 289 #molecular-weight 33514 #checksum 4703

```

Query Match	7.2%	Score 167;	DB 11;	Length 289;
Best Local Similarity	28.6%	Pred. No. 1.97e-09;		
Matches	60;	Conservative	46;	Mismatches 91; Indels 13; Gaps 11;
Db	40	pferqfyindltishpyatterrvnmmlfyfsvpsltillig--sl--ledrrhlifi	95	
Qy	61	PYHRGFYNDSIKYPKTGTTINDAVLCVAGVIAITAIETGEFFRIYILKRSRTIQN	120	
Db	96	lytsllglsla-wfst-s-fftnfkhwgrlpdpfldrcqp-veglpdlftfakdv	149	
Qy	121	PYVAALYQVCCFLFGCAISQSESTDAKVSIGRLPHEFLSVCPDFOINCSEGYQNYR	180	
Db	150	cttkaherllgffttspgnssesfaglylwlcgqlltesplmpwkrkmvafipllg	209	
Qy	181	CRG-DDSKVQEARKSFFSGHASFMTMLYLVLYLQARETWGARL-L-RPLLOFTLIMM	237	
Db	210	aalialsrstqdyrrhfdvllgsmlygima	239	
Qy	238	AFVTGLLSRVSDHKHHPSPDVLAGFAQALVA	267	

```

RESULT      3
ENTRY
TITLE       S32217      #type complete
            hypothetical protein 2 - Bacillus megaterium
FORMAL_NAME #formal_name Bacillus megaterium
DATE        06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
            06-Feb-1995
ACCESSIONS  S32217
REFERENCE   S32215
AUTHORS     Rauschenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.;
            Boidol, W.; Siewert, G.
SUBMISSION  submitted to the EMBL Data Library, March 1993
DESCRIPTION Cloning, sequencing and expression of the gene for cytochrome
            P450me6, the steroid 15beta-monooxygenase from Bacillus
            megaterium ATCC 13368.
ACCESSION   S32217
STATUS      preliminary
MOLECULE    #molecule_type DNA
RESIDUES    #residues 1-216 ##label RAU
CROSS-REFS  #cross-references EMBL: Z21972
SUMMARY     #length 216
            #object_id 24046 #object_name 9213

```

Query Match	5.9%	Score 138;	DB 10;	Length 216;
Best Local Similarity	38.6%;	Pred. No. 2.46e-05;		
Matches	32; Conservative	20; Mismatches 23;	Indels 8;	Gaps 5;
Db	127	sfpsghammafslgylftllwrlhtarw--axil--llfsmllmi-lslglsrlylgv	181	
		:    :   :    :    :    :    :    :		
QY	194	SFSFGA--SFSMTMLYLIVLYLQARTWRGALLRPLLQFTLIMMAFYTGLSRVSDHKH	251	
		:    :    :    :    :    :    :		
Db	182	ypsddliagylagcgwlnslwff	204	
		:    :    :    :		
QY	252	HPSDVLVLAGFAOGA-LVACCIVFF	273	

RESULT	4
ENTRY	S58067
TITLE	#type fragment
ORGANISM	probable olfactory receptor tpcr34 - mouse (fragment)
DATE	#formal_name Mus musculus #common_name house mouse
	13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change

```

19-Apr-1996
#accessions S58067
#reference S57995
#authors vanderhaeghen
#submission submitted to 1
#description Male germ cell
specific recombination
#accession S58067
#status preliminary
#molecule_type mRNA
#features 1-157 ##
#residues 157
#cross-references EMBL:1
#length 157
SUMMARY

```

```

Query Match          5.5%; Score 128; DB 14; Length 157;
Best Local Similarity 27.6%; Pred. NO. 5.33e-04;
Matches 27; Conservative 30; Mismatches 33; Indels 8; Gaps 8;

Db 18 llvflsvlsifvseqllhnsvwldtyfkfadvishdfcdpsql-inlacsdtftnmvmyf 76
QY 35 LLTICLD-LFCFLMAGLP-FLLIETSKYFHRG-FYCNDESIKYLKGTETI-NDAVLCA 90

Db 77 vqtisgfpip-sgaifs-yy-kiavsaillmpagpgkyk 111
QY 91 VGVITAILAIIITGFEYRIYVYLKKSRTSTPNPVVAALYK 128

```

RESULT	5
ENTRY	S28996
TITLE	G protein-coupled receptor (clone PTE01) - rat (fragment)
ORGANISM	#formal_name Rattus norvegicus #common_name Norway rat
DATE	22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995

```

ACCESSIONS
REFERENCE
#authors Abe, K.; Kusakabe, Y.; Tanemura, K.; Emori, Y.; Arai, S.
#journal FEBS Lett. (1993) 316:253-256
#title Multiple genes for G protein-coupled receptors and their
        expression in lingual epithelia.
#accession S28996
        #molecule_type mRNA
        #residues 1-185 #label ABE
SUMMARY #length 185 #checksum 2049

```

	Query Match	5.2%; Score 122; DB 14; Length 185;
	Best Local Similarity	28.6%; Pred. NO. 3.20e-03;
	Matches	28; Conservative 30; Mismatches 32; Indels 8; Gaps 8;
Db	85 llvflsilslvlsvqlnsvldltqyfkavdshlffc-dpalllnalcsdtftnnlmvyf	l13
	:    :    :    :    :    :    :	:   :   :   :
QY	35 LLICLD-LFCFLMAGLR-FLLIETSIKPYHRG-EYCNDSEIKYPLKTGETI-NDAVLCA	90
Db	144 vgaigsgfli-sgiffs-yy-kivssllmpsgggkyk	178
	:    :    :    :    :    :	:   :   :   :
QY	91 VGIIVTALLAITGGFYRYLYLKSRGTIONPVVAALYK	128
	:    :    :    :    :    :	:   :   :   :

```

6
RESULT      6
ENTRY       S50669      #type complete
TITLE       hypothetical protein YER166w - yeast (Saccharomyces
            cerevisiae)
ORGANISM    #formal_name Saccharomyces cerevisiae
            28-May-1993  #sequence_revision 24-Feb-1995  #text_change
            03-May-1996
DATE
ACCESSIONS S50669; S30822
REFERENCE   S50428;
            Dietrich, F.S.
            submitted to the EMBL Data Library, December 1994
            The sequence of S. cerevisiae cosmids 9163 and 9132.
            S50669
            #molecule_type DNA
            #residues      1-1571  ##label DIE
            ##cross-references EMBL:U18922

```

```

SEQUENCE
S03812
Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.;
Komp, C.; Wei, Y.; Taylor, P.; Nakahara, K.; Roberts, D.;
Davis, R.W.
#submission submitted to the EMBL Data Library, February 1993
#accession S03822
##molecule_type DNA
##residues 1-222 ##label MUL
##cross-references EMBL:L0718
GENETICS
#map_position 5R
SUMMARY
length 1571 #molecular-weight 177797 #checksum 2315
Query Match 5.2%; Score 120; DB 11; Length 1571;
Best Local Similarity 20.0%; Pred. No. 5.75e-03;
Matches 18; Conservative 29; Mismatches 40; Indels 3; Gaps 3;
Db 1244 vndtlslvpqlrvglrrtewqrkflwymldgylqslciffpylvyhknmtvgl 1303
| : :::: :||: ||: | : || : ||: | : : : : : : : : : : : :
QY 94 VTALAITGFERYIKKSRSTIQNP-YV-AALYKQGCVLFCAISQSFDTIAKVISI 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1304 gidhyrfygyvvtiaivscntyllhqyr 1333
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 GLRLPHELSVCNPDFSQINCSG-CYIQNYR 180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 7
ENTRY #type complete
TITLE probable membrane protein YGR036c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein G4085
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Jul-1996
ACCESSIONS S64327
REFERENCE S64071
#authors Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64327
##molecule_type DNA
##residues 1-239 ##label RIE
##cross-references EMBL:D72821
##experimental_source strain S288C
GENETICS
#map_position 7R
KEYWORDS Transmembrane protein
FEATURE
39-55 #domain transmembrane #status predicted #label TM1\
136-152 #domain transmembrane #status predicted #label TM2\
169-185 #domain transmembrane #status predicted #label TM3
SUMMARY
length 239 #molecular-weight 27649 #checksum 3184
Query Match 5.1%; Score 118; DB 12; Length 239;
Best Local Similarity 27.4%; Pred. No. 1.03e-02;
Matches 17; Conservative 21; Mismatches 22; Indels 2; Gaps 2;
Db 126 swknlfielkfsgatallsfecvfervlyhynldqvlgfsvgalts-lyffivgi 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 TWRGALL-RPLQLFTLLMAFYGLSRVSDKHHPDVLAFAGALIVACCIVFFVS DL 277
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 lr 186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 FK 279
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 8
ENTRY #type fragment
TITLE G-protein coupled receptor Gpcrl3 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 01-Dec-1995
ACCESSIONS E48909
REFERENCE E48909
```

```

#authors      Wilkie, T.M.; Chen, Y.; Gilbert, D.J.; Moore, K.J.; Yu, L.;
#journal      Simon, M.I.; Copeland, N.G.; Jenkins, N.A.
#journal      Genomics (1993) 18:175-184
#title        Identification, chromosomal location, and genome organization
              of mammalian G-protein-coupled receptors.
#accession    E48909
#status       preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues     1-180 ##label WIL
#cross-references GB:L20334
SUMMARY       #length 180 #checksum 3623

Query Match      4.8%; Score 111; DB 14; Length 180;
Best Local Similarity 37.8%; Pred. No. 7.57e-02;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

Db 130 vlcvrtfsvillaivalvriyfvrrshadvgpqtllalkvcv 174
      ||| | | | | | | | | | | | | | | | | | | |
QY 87 VLCAGVIGTAVILAIIITGEFYRIYL-KKSRSTIQNPVVAALYKQV 130

RESULT 9
ENTRY   JG1465 #type complete
TITLE   Probable G protein-coupled receptor - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE    30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
        13-Sep-1996
ACCESSIONS JG1465
REFERENCE   JG1465
#authors    Okazaki, H.; Ishizaka, N.; Sakurai, T.; Kurokawa, K.; Goto,
            K.; Kumada, M.; Takawa, Y.
#journal    Biochem. Biophys. Res. Commun. (1993) 190:1104-1109
#title      Molecular cloning of a novel putative G protein-coupled
            receptor expressed in the cardiovascular system.
#accession  JG1465
#molecule_type mRNA
#residues   1-352 ##label OKA
#experimental_source aortic smooth muscle
KEYWORDS    glycoprotein; phosphoprotein; transmembrane protein
FEATURE     35-59
            #domain transmembrane #status predicted #label TM1\
            #domain transmembrane #status predicted #label TM2\
            #domain transmembrane #status predicted #label TM3\
            #domain transmembrane #status predicted #label TM4\
            #domain transmembrane #status predicted #label TM5\
            #domain transmembrane #status predicted #label TM6\
            #domain transmembrane #status predicted #label TM7\
            #binding_site carbohydrate (Asn) (covalent) #status
            predicted\
142,145,218,219,
329,330,331,332 #binding_site phosphate (Ser) (covalent) #status
            predicted\
313 #binding_site phosphate (Thr) (covalent) #status
            predicted
SUMMARY     #length 352 #molecular-weight 38734 #checksum 3672

Query Match      4.8%; Score 111; DB 14; Length 352;
Best Local Similarity 37.8%; Pred. No. 7.57e-02;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

Db 191 vlcvrtfsvillaivalvriyfvrrshadvgpqtllalkvcv 235
      ||| | | | | | | | | | | | | | | | | | | |
QY 87 VLCAGVIGTAVILAIIITGEFYRIYL-KKSRSTIQNPVVAALYKQV 130

RESULT 10
ENTRY   S26033 #type complete
TITLE   NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 -
            Caenorhabditis elegans mitochondrion (SGC4)
            #formal_name mitochondrion Caenorhabditis elegans
            12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
            09-Sep-1994
ACCESSIONS S26033; S25805

```

```

REFERENCE
#authors      S26014
               Okimoto, R.; Macfarlane, J.L.; Clary, D.O.; Wolstenholme,
               D.R.
#journal      Genetics (1992) 130:471-498
#title        The mitochondrial genomes of two nematodes, Caenorhabditis
               elegans and Ascaris suum.
#cross-references MUID:92201635
#accession    S26033
               #molecule_type DNA
               #residues_type 1-409 #label OKI
               #cross-references EMBL:X54252
               #note the authors translated the initiation codon TTG for
               residue 1 as Leu

REFERENCE
#authors      S13139
               Okimoto, R.; Macfarlane, J.L.; Wolstenholme, D.R.
#journal      Nucleic Acids Res. (1990) 18:6113-6118
#title        Evidence for the frequent use of TTG as the translation
               initiation codon of mitochondrial protein genes in the
               nematodes, Ascaris suum and Caenorhabditis elegans.
#accession    S25806
               #molecule_type DNA
               #residues_type 1-25 #label OK2
               #cross-references EMBL:X54252
               #note the authors translated the initiation codon TTG for
               residue 1 as Leu

GENETICS
#gene         ND4
#genome       mitochondrion
#genetic_code SGC4
#start_codon  TTG
CLASSIFICATION
#superfamily  NADH dehydrogenase (ubiquinone) chain 4
#membrane-associated complex; mitochondrion; NAD: oxidative
#phosphorylation; oxidoreductase; respiratory chain
#length 409 #molecular-weight 47206 #checksum 6596

SUMMARY
Query Match 4.8%; Score 112; DB 5; Length 409;
Best Local Similarity 27.6%; Pred. No. 5.72e-02;
Matches 21; Conservative 25; Mismatches 22; Indels 8; Gaps 7;

Db 18 flftvmfslifn--fswggflvldsysfillvmslilglilvise-knn--nlli- 72
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 202 FSNYIMLVLYLVLOARFTWRCARLLRPLLOFTL-IMAAFT-GLSRVSDKHPDSVLAG 259
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 73 lse-llvfcliffip 87
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 260 FAQGALVACCVFFVS 275

RESULT 11
ENTRY
TITLE        #type fragment
ORGANISM      probable olfactory receptor tpcr62 - dog (fragment)
#formal_name  Canis lupus familiaris #common_name dog
DATE          13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
01-Mar-1996
ACCESSIONS    S57995
REFERENCE      S57995
#authors      Vanderhaeghen, P.; Schurmans, S.; Vassart, G.; Parmentier, M.
#submission   submitted to the EMBL Data Library, July 1995
#description   Male germ cells from several mammalian species express a
               specific repertoire of olfactory receptor genes.
#accession    S57995
               #status preliminary
               #molecule_type mRNA
               #residues 1-157 #label VAN
               #cross-references EMBL:X89660
               #length 157 #checksum 221

SUMMARY
Query Match 4.7%; Score 109; DB 13; Length 157;
Best Local Similarity 25.3%; Pred. No. 1.32e-01;
Matches 23; Conservative 29; Mismatches 32; Indels 7; Gaps 7;

Db 25 lislsvqmhslvqltyfkdvslshffcdpsql-lnlacsdtdfinnlvmvfvgaisgf 83
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

QY 41 LFCLEMACLP-FLIERTSTIKPYHRG-FYCNDSEIKYPLKTGET-INDAVLCAGVIVIAI 97

Db 84 lpi-sgiffs-yy-kivssilrvpstggyrk 111
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 98 LAITGFEYRIYYLKKSRSTIONPYVAALYK 128

RESULT 12
ENTRY
TITLE        #type complete
ORGANISM      MHC class II alpha chain - Danio rerio
#formal_name  Danio rerio
DATE          13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS    I51740
REFERENCE      I51740
#authors      Sultmann, H.; Mayer, W.E.; Figueroa, F.; O'Huigin, C.; Klein,
               J.
#journal      Immunogenetics (1993) 38:408-420
#title        Zebrafish MHC class II alpha chain-encoding genes:
               polymorphism, expression, and function.
#cross-references MUID:94011091
#accession    I51740
               #status preliminary
               #residues 1-236 #label SUL
               #cross-references GB:L19445; NID:g311192; CDS_PID:g311193
               #length 236 #molecular-weight 26136 #checksum 3945

SUMMARY
Query Match 4.3%; Score 99; DB 16; Length 236;
Best Local Similarity 37.0%; Pred. No. 1.94e-00;
Matches 10; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

Db 204 svgpavfcgvgilvgllgvaagtffli 230
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 82 TINDAVLCAGVIVIAITAITGFEYRI 108

RESULT 13
ENTRY
TITLE        #type complete
ORGANISM      MHC class II alpha chain - Danio rerio
#formal_name  Danio rerio
DATE          13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS    I51745
REFERENCE      I51745
#authors      Sultmann, H.; Mayer, W.E.; Figueroa, F.; O'Huigin, C.; Klein,
               J.
#journal      Immunogenetics (1993) 38:408-420
#title        Zebrafish MHC class II alpha chain-encoding genes:
               polymorphism, expression, and function.
#cross-references MUID:94011091
#accession    I51745
               #status preliminary
               #residues 1-236 #label SUL
               #cross-references GB:L19450; NID:g311202; CDS_PID:g311203
               #length 236 #molecular-weight 26205 #checksum 7205

SUMMARY
Query Match 4.3%; Score 99; DB 16; Length 236;
Best Local Similarity 37.0%; Pred. No. 1.94e-00;
Matches 10; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

Db 204 svgpavfcgvgilvgllgvaagtffli 230
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 82 TINDAVLCAGVIVIAITAITGFEYRI 108

RESULT 14
ENTRY
TITLE        #type complete
ORGANISM      MHC class II alpha chain - Danio rerio
#formal_name  Danio rerio
DATE          13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
13-Sep-1996
ACCESSIONS    I51741
REFERENCE      I51741

```



Search completed: Tue Nov 4 10:27:30 1997  
Job time : 82 secs.

**This Page Blank (uspto)**

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 11:36:00 1997; Maspar time 1135.34 Seconds  
Tabular output not generated.  
1317.542 Million cell updates/sec

Title: >US-08-842-827-5  
Description: (1-1362) from US08842827.seq  
Perfect Score: 1362  
N.A. Sequence: 1 GGCGCAGCTGCAAGATT.....TTTTAAAAA.....AAAAA 1362  
Comp: CCGCGTGAGACGTTTCAA.....AAAAATTTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb1-new3  
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database: genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PLN12 71: PLN13  
72: PLN14 73: PLN15 74: PLN16 75: PLN17 76: PLN18 77: PLN19  
78: PLN20 79: PLN21 80: PLN22 81: PLN23 82: PLN24 83: PLN25 84: PLN26  
85: PLN27 86: PLN28 87: PLN29 88: PLN30 89: PLN31 90: PLN32 91: PLN33  
92: PLN34 93: PLN35 94: PLN36 95: PLN37 96: PLN38 97: PLN39 98: PLN40  
99: PLN41 100: PLN42 101: PLN43 102: PLN44 103: PLN45 104: PLN46  
105: PLN47 106: PLN48 107: PLN49 108: PLN50 109: PLN51 110: PLN52  
111: PLN53 112: PLN54 113: PLN55 114: PLN56 115: PLN57 116: PLN58 117: PLN59  
118: PLN60 119: SYN 120: UNA 121: VRL  
u-emb150\_99  
122: part1

Database: genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL

Database: u-emb150\_99  
122: part1

Statistics: Mean 11.681; Variance 5.524; scale 2.114

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1131	83.0	1444	HSU79294	Human clone 23748 mRN	0.00e+00
2	784	57.6	2206	91 RNDRI42	R.norvegicus mRNA for	0.00e+00
3	102	7.5	1212	85 D84376	House mouse; Musculus	1.50e-56
4	85	6.2	1622	40 DMU73822	Drosophila melanogast	8.36e-43
5	71	5.2	1490	88 MUSHPIP	Mus musculus (clone H	8.69e-32
6	35	2.6	215	57 128278	Sequence 5 from paten	6.42e-06
7	33	2.4	215	57 128278	Sequence 5 from paten	1.12e-04
8	28	2.1	1185	90 RATCERAC	Rat Cardioembryonic	9.77e-02
9	27	2.0	3203	79 HOMBMLIX	Human prot-oncogene (	3.50e-01
10	27	2.0	4204	91 RAINUCIA3	Rat nucleolin gene.	3.50e-01
11	26	1.9	4251	88 MUSA55RNA	Mouse 4.5S RNA gene.	1.22e+00
12	26	1.9	38118	33 CET27E9	*** SEQUENCING IN PRO	1.22e+00
13	26	1.9	58382	110 CET28A8	Caenorhabditis elegans	1.22e+00
14	26	1.9	58382	5 CET28A8	Caenorhabditis elegans	1.22e+00
15	26	1.9	170789	110 CEY76A2	Caenorhabditis elegans	1.22e+00
16	26	1.9	170789	5 CEY76A2	Caenorhabditis elegans	1.22e+00
17	25	1.8	99	55 113410	Sequence 12 from pate	4.09e+00
18	24	1.8	201	53 A10161	Synthetic DNA for pre	1.33e+01
19	24	1.8	201	53 A10159	S.griseus gene for pr	1.33e+01
20	25	1.8	801	50 GGRSFR	G.gallus RSFR mRNA.	4.09e+00
21	24	1.8	870	85 MMHGM2	M.musculus mRNA for h	1.33e+01
22	24	1.8	921	122 S78239	BSP-bone sialoprotein	1.33e+01
23	24	1.8	1400	13 RGU67155	Rubrivivax gelatinosu	1.33e+01
24	24	1.8	1770	38 CSU43316	Caenorhabditis elegans	1.33e+01
25	24	1.8	1914	69 YSPORF3	Schistosaccharomyces p	1.33e+01
26	24	1.8	2170	82 HUMOCT6A	Homo sapiens (oct-6)	1.33e+01
27	24	1.8	2502	122 HSD347	Human mRNA for Neurod	1.33e+01
28	24	1.8	2502	117 D82347	Human mRNA for Neurod	1.33e+01
29	24	1.8	2578	50 GGU67889	Gallus gallus bone si	1.33e+01
30	25	1.8	2708	41 DROZSTA	S.cerevisiae chromoso	4.09e+00
31	24	1.8	2760	67 SCYOR382W	Mouse zic mRNA for zi	1.33e+01
32	24	1.8	2947	14 MMZIC	Mouse mRNA for zic pr	1.33e+01
33	24	1.8	2947	90 MOSZIC	Human acid finger pro	1.33e+01
34	24	1.8	3595	77 HSU009825	Human calcium depende	1.33e+01
35	24	1.8	3769	77 HSU11717	Human calcium depende	1.33e+01
36	24	1.8	3987	77 HSU11058	Adeno-associated viru	1.33e+01
37	24	1.8	4726	96 AVU48704	Artificial sequences	1.33e+01
38	24	1.8	6789	94 ASPDXD3C	Drosophila virilis mo	1.33e+01
39	24	1.8	7154	41 DROGYPE5Y2	Caenorhabditis elegans	1.33e+01
40	24	1.8	27220	35 CEC3487	Caenorhabditis elegans	1.33e+01
41	25	1.8	34028	37 CELF59A6	Human DNA from cosmid	4.09e+00
42	25	1.8	37940	79 HUM19ANON1	Caenorhabditis elegans	4.09e+00
43	25	1.8	39320	36 CEF13E6	Equine herpesvirus 2,	1.33e+01
44	24	1.8	184427	97 EHVU20824	Human DNA sequence **	1.33e+01
45	24	1.8	195158	34 HS473B4		

## ALIGNMENTS

RESULT 1  
LOCUS HSU79294 1444 bp mRNA PRI 30-JAN-1997  
DEFINITION Human clone 23748 mRNA, complete cds.  
ACCESSION U79294  
NID g1710275

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1444)

AUTHORS Anderson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.

TITLE A 'double adaptor' method for improved shotgun library construction

JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)

MEDLINE 96207227

REFERENCE 2 (bases 1 to 1444)

AUTHORS Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,

Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.

TITLE Large Scale Concatenation cDNA Sequencing

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1444)

AUTHORS Yu, W. and Gibbs, R. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-NOV-1996) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA  
 FEATURES Location/Qualifiers  
 source 1..1444  
 /organism="Homo sapiens"  
 /clone="23748"  
 /sex="female"  
 /dev\_stage="infant"  
 /tissue\_type="brain"  
 /clone\_lib="Soares library lNIB from IMAGE consortium"  
 375..1076  
 /codon\_start=1  
 /product="unknown"  
 /db\_xref="PID:g1710275"  
 /translation="MONKYDKAIVPESKNGSPALNNPRSGKRVLLICLDLFL  
 FWAGLPFLIETSTKPVHRGFCYNDESIKYPLKGTETINDAVLCVAVIALLIIT  
 GEYRIYIKKSKSTIQNPYAAIKQVCFGLCAISQSFIDIAKVSIRLPHFLS  
 VCNPFDSQINCSEGIYONRCRDDSKVQEARKSFFSGHASFMYTLYLVLYQARE  
 TWGAPCSGSPSSP"  
 BASE COUNT 327 a 427 c 373 g 317 t  
 ORIGIN

Query Match 83.08; Score 1131; DB 78; Length 1444;  
 Best Local Similarity 99.98; Pred. No. 0.00e+00;  
 Matches 1137; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 306 gcgcctgggtgtggtgctgtgtgctgagcgtcttcggtggcgaggtcgcgcgcga 365  
 QY 225 CGCCTGGGTGTGCTGCTGTGTGGCGGACGCTTCGCGGGCGGGAGGCTCGCGCCGCA 284

Db 366 gccagcggcgcacaaactcaagtcagcaagaagatcgctcccgagagcaagaagcgc 425  
 QY 285 GCCAGCGCATGCAAAACTACAAGTACGACAAAGCATCGTCCCGGAGAGCAAGACGGC 344

Db 426 ggcagcggcgctccaaacaacacggagagagcgagcaagcggtgctctcctc 485  
 QY 345 GCGACGCCGGCGCTCAACAACAACCGGAGGAGGAGCGGAGCGGAGGCTGCTCATC 404

Db 486 tgcctcgacctcttcgctcttcacgagcgccctccctctcctcctcctcagcaagc 545  
 QY 405 TGCTCGACCTCTCTGCTCTCTCATGTGGGGCGCTCCCTCTCTCATCTGACACAGC 464

Db 546 accatcaagccttaccacgggggtttactgcaatgatgagagcatcaagtaccactg 605  
 QY 465 ACCATCAAGCCTTACCACCGAGGGTTTACTGCAATGATGAGAGCATCAAGTACCACATG 524

Db 606 aaactggtgagacaataaatagcgtgtgctctgtgctgtgctgtggtggtcgtcgtcc 665  
 QY 525 AAACCTGGTGAGACAATAATGACGTGTGCTGTGCGGTGGGATCGTCAATGCCATC 584

Db 666 ctccgcatcatcagcgggggaattctaccggatctattaccatgaagaagtcggtgcagc 725  
 QY 585 CTCGCGATCATCAGCGGGGAATTCTACCGGATCTATTACCTGAAGAAGTCGCGTCCAGC 644

Db 726 attcagaacctcagtcggcagcactctataagcaagtggctgtctcctcttggctgt 785  
 QY 645 ATTCAGAACCCCTACGTGGGAGCACTCTATAGCAAGTGGGCTGCTCTCTTTGGGTGT 704

Db 786 gccatcgcagcagcttttcacagacatgcacaaagtgtccatggcgctcgctcctcac 845  
 QY 705 GCCATCAGCCAGCTTTTCACAGACATTTGCCAAGTGTCCATAGGCGCGCTCGCTCCAC 764

Db 846 ttcttgagtgctgcacccctgatttcacagacataactgctctggaaggtcatattcag 905  
 QY 765 TTCITGAGTGCTCTCAACCCCTGATTTCAGCCAGATCACTGCTCTGAAGGCTACATTCAG 824

Db 906 aactacagatgcagagtgatgcagcaagaagtcacgaagccaggaagtcctcttctct 965  
 QY 825 AACTACAGATGCAGAGTGATGACAGCAAGTCCAGGAAGCCAGGAAGTCCTCTTCTCT 884

Db 966 ggcacatgcctctctccatgtacactatgctgtatttgggtgctatacctgcagggccgcg 1025

QY 885 GGCCATGCCCTCCCTTCCTCCATGTACACTATGCTGTATTGGTGCTATACCTGCGAGGCCGC 944  
 Db 1026 ttcaacttggcagagagccgcgc-tgctccggccccctccctgcagttccacttgcacatgatg 1084  
 QY 945 TTCACTTGGCAGGAGCCCGCTGCTCCGGCCCTCTCTGCGAGTTCCACTTGTATCATGATG 1004  
 Db 1085 gcctttcacagggagactgctcgcgtatcagaccacaagcaccatccagtgatgtctc 1144  
 QY 1005 GCCTTTCACACGGGACTGTCTCGGTATCAGACCACAAGCACCATCCCATGTGTTCTG 1064  
 Db 1145 gcaggatttgcacagagccctgggtggcgtgctgcacagtttcttctgctgctgacctc 1204  
 QY 1065 GCAGGATTTGCTCAAGGAGCCCTGGTGGCCTGCTGCATAGTTCCTTCGTTGCTGACCTC 1124  
 Db 1205 ttcaagactaagcagcgtctctccctgctgctccctgctatcccgaaaggaatcccttca 1264  
 QY 1125 TTAAGACTTAAGACGAGCGCTCTCCCTGCTGCTGCCCTGTATCCGGAAGAAATCCCTTTCA 1184  
 Db 1265 cctgtgacacattattacacaggaacaatcaccaacacatgatgtaggcgccaccacctcc 1324  
 QY 1185 CCTGTGCACATTTATTGACAGGAACAATCAGCACACATGATGATGAGTGCCACCCACCTCC 1244

Db 1325 tgagcttttttataaaatgactgctgacagcaagttcttctgctctcctcaatcctca 1384  
 QY 1245 TGAGCTGTTTTTAAATGACTGCTGACAGCAAGTTCTTCTGCTCTCTCAATCTCATCA 1304

Db 1385 gacagtagaagttagggaaaacttttggccgactgatttttaaaaaaataaaaaa 1442  
 QY 1305 GACAGTAGAATGTAGGAAAACCTTTGCGCGACTGATTTTAAAAAATAAAAAA 1362

RESULT 2  
 LOCUS R.ndr142 2206 bp RNA ROD 22-NOV-1996  
 DEFINITION R.norvegicus mRNA for ER transmembrane protein.  
 ACCESSION Y07783  
 NID gi684744  
 KEYWORDS Dri 42 gene; ER-transmembrane protein.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 2206)  
 AUTHORS Barila, D., Plateroti, M., Nobili, F., Muda, A.O., Xie, Y., Morimoto, T. and Perozzi, G.  
 TITLE The Dri 42 gene, whose expression is up-regulated during epithelial differentiation, encodes a novel endoplasmic reticulum resident transmembrane protein  
 JOURNAL J. Biol. Chem. 271 (47), 29928-29936 (1996)  
 MEDLINE 97094703  
 REFERENCE 2 (bases 1 to 2206)  
 AUTHORS Perozzi, G.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-SEP-1996) G. Perozzi, Istituto Nazionale Della Nutrizione, Unit Of Experimental Nutrition, Via Ardeatina 546, 00178 Roma, ITALY  
 FEATURES Location/Qualifiers  
 1..2206  
 /organism="Rattus norvegicus"  
 /strain="Wistar"  
 /cell\_type="epithelial"  
 /dev\_stage="adult"  
 /tissue\_type="small intestine"  
 410..1348  
 /gene="Dri 42"  
 /codon\_start=1  
 /product="ER transmembrane protein"  
 /db\_xref="PID:e283078"  
 /db\_xref="PID:g1684745"  
 /translation="MQSYKDYKAIVPESKNGSPALNNPRSGKRVLLICLDLFL  
 FNAALPFLIETSTKPVHRGFCYNDESIKYPLKVSITINDAVLCVAVIALLIIT  
 GEYRIYIKKSKRSTIQNPYAAIKQVCFGLGCAISQSFIDIAKVSIRLPHFL  
 CDS



[illegible]

```

LOCUS       128278       215 bp      DNA              PAT              30-OCT-1996
DEFINITION   Sequence 5 from patent US 5569830.
ACCESSION   128278
NID         91819054
KEYWORDS     .
SOURCE      Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 215)
AUTHORS     Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE       Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
JOURNAL     Patent: US 5569830-A 5 29-OCT-1996;

FEATURES             Location/Qualifiers
     source           1..215
     BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN
Query Match      2.68; Score 35; DB 57; Length 215;
Best Local Similarity 13.5%; Pred.No. 6.42e-06;
Matches 28; Conservative 86; Mismatches 91; Indels 3; Gaps 3;

Db 6 msssvvrtascndkakkdgnitsswttdccnrtwgvcdtdttrvnn-dsgnkhkyssa 64
Cp 1092 CCACGAGGCTCTTGCAAAATCTGCCAGAACATCATCTGGATGCTGTGGTCTG 1033
Db 65 nnyggnnvgaakthyythnvsgadskvtvdsynasgtssnggt-dgnrsgadsvgss 123
Cp 1032 ATAGCGAGACAGTCCCGTGTAGAGGGCCATCATCAAGTGAACCTGCAGGAGGGGCC 973
Db 124 kta-mtsnrnrtgtannavdsrnmgdasvgsdkntkhhaknsadgkvgsknngdrnnryg 182
Cp 972 GGAGCAGCGGGCTCTCCCAAGTGAAGCGGCTCGAGGTATAGCACCAATACAGCA 913
Db 183 tgtknsvnncgggnkrdvssyannkcc 210
Cp 912 TAGTGTACATGGAGAGGAGGATGCC 885

RESULT 7
LOCUS       128278       215 bp      DNA              PAT              30-OCT-1996
DEFINITION   Sequence 5 from patent US 5569830.
ACCESSION   128278
NID         91819054
KEYWORDS     .
SOURCE      Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 215)
AUTHORS     Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
TITLE       Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
JOURNAL     Patent: US 5569830-A 5 29-OCT-1996;

FEATURES             Location/Qualifiers
     source           1..215
     BASE COUNT      15 a      8 c      25 g      26 t      141 others
ORIGIN

Query Match      2.48; Score 33; DB 57; Length 215;
Best Local Similarity 10.5%; Pred.No. 1.12e-04;
Matches 21; Conservative 84; Mismatches 94; Indels 1; Gaps 1;

Db 6 msssvvrtascndkakkdgnitsswttdccnrtwgvcdtdttrvnn-dsgnkhkyssan 65
Qy 86 AGAGCCCGGAGGAGGAGCAAAACCCACTTCTGCAGTGTGGAGGAGCAGCAGCCCGG 145
Db 66 ynyggnnvgaakthyythnvsgadskvtvdsynasgtssnggt-dgnrsgadsvgsskt 125
Qy 146 CTGACACTTAGCCGCGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 205

```

```

Db 126 amtsnrnrtgtannavdsrnmgdasvgsdkntkhhaknsadgkvgsknngdrnnrygtgt 185
Qy 206 GCCCAGGAGGCGACCGCGGCGCTGTGGTGTGTGGCTGCTGTGGCGGACGCTTT-CGCGG 264
Db 186 ksnvsnncgggnkrdvssya 205
Qy 265 GCGCGGAGGCGTCGCGCGCA 284

RESULT 8
LOCUS       RATCEAC      1185 bp      DNA              ROD              16-JAN-1991
DEFINITION   Rat carcinoembryonic antigen (CEA3) gene, exon X.
ACCESSION   M60025 J04626 M2228
NID         9203405
KEYWORDS     carcinoembryonic antigen.
SOURCE      Rat (strain BD II) liver DNA, clone rncGM3.
ORGANISM     Rattus norvegicus
Eukaryotae; mitochondria; eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Rattus.
REFERENCE    1 (bases 1 to 1185)
AUTHORS     Kodell,J.V., Lucas,K., Barnert,S., von Kleist,S., Thompson,J.A. and
            Zimmermann,W.A.
TITLE       Identification of a carcinoembryonic antigen gene family in the
            rat: Analysis of the N-terminal domains reveals
            immunoglobulin-like, hypervariable regions
JOURNAL     J. Biol. Chem. 264, 6906-6912 (1989)
MEDLINE     89214106
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            W.Zimmermann 12-JAN-1989.

FEATURES             Location/Qualifiers
     source           1..1185
     /organism="Rattus norvegicus"
     /strain="BD II"
     /tissue_type="liver"
     /partial
     intron            1..728
     repeat_unit       /gene="CEA3"
                        138..203
     repeat_unit       /gene="CEA3"
                        412..514
     CDS               /gene="CEA3"
                        739..1088
                        /partial
                        /gene="CEA3"
                        /codon_start=3
                        /product="carcinoembryonic antigen"
                        /db_xref="PID:g594436"
     /translation="SLITICMLLPHTAHVTLKSSPPQVVEGENVLLSAOHLPENIIAFA
            WTKGTDNRGIALYSRTVSLTGPVHSGRETLYSDGSLWKNVTDGTGFTFRIL
            NNHGKIQSNITFLHVK"
     intron            1089..1185
                        /partial
                        /gene="CEA3"

BASE COUNT      312 a      279 c      277 g      317 t
ORIGIN

Query Match      2.18; Score 28; DB 90; Length 1185;
Best Local Similarity 75.0%; Pred.No. 9.77e-02;
Matches 42; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 410 atgaggagaggaacgggaagaggaggaggaggaggaggaggaggaggaggag 465
Cp 451 ATGAGGAAGGGAGGCCGCCCATGAAGAGGAGGAGAGAGGTCGAGGAGATGACGAG 396

RESULT 9
LOCUS       HUMBMILX     3203 bp      mRNA              PRI              19-JUL-1994
DEFINITION   Human prot-oncogene (BMI-1) mRNA, complete cds.
ACCESSION   L13689
NID         9291872
KEYWORDS     prot-oncogene.

```

```

SOURCE      Homo sapiens cDNA to mRNA..
ORGANISM    Homo sapiens
REFERENCE    1 (bases 1 to 3203)
AUTHORS     Alkema,M.J., Wiegant,J., Raap,A.K., Berns,A. and van Lohuizen,M.
TITLE       Characterization and chromosomal localization of the human
            proto-oncogene BMI-1
JOURNAL     Hum. Mol. Genet. 2 (10), 1597-1603 (1993)
MEDLINE     94093345
FEATURES    Location/Qualifiers
            1..3203
            /organism="Homo sapiens"
            /cell_line="K562"
            /germline
            1..3203
            /gene="BMI-1"
            1..479
            /gene="BMI-1"
            /note="putative"
            480..1460
            /gene="BMI-1"
            /note="putative"
            /codon_start=1
            /function="Prot-oncogene"
            /db_xref="PID:g291873"
            /translation="MHRTRIKITELNPLHMCVLGGYFIDATTIECLHSFCKTCIV
            RYLETSCYICDVQVHKTRPLINIRSDKLTQDIVKLVPLGLENKRRDRFYAAHP
            SADAANGSNDGRGVADEKRIITDDEIISIEFFDONRLDRKVKDKKSEEVND
            KRYLRCAATVMHLRAKELRSKMDIPNTFQIDVMEPLKDYTLMDIAYIVTWRRN
            GPLPLKYVRPTCKRMKISHORDGLTNAGELESDSGSKANSPAGGVPSTSSCLPSPS
            TPVQSPHQPFPHTSSMTNGTSNPSGNHQSSEFANRPRKSSVNGSSATSSG"
            1461..3203
            /gene="BMI-1"
            /evidence=experimental
            3203
            /gene="BMI-1"
            /polyA_site
BASE COUNT  959 a 591 c 678 g 975 t
ORIGIN
Query Match      2.0%; Score 27; DB 79; Length 3203;
Best Local Similarity 72.1%; Pred. No. 3.50e-01;
Matches 44; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Db 154 cgaggagagaggagggagggccggaggagggcgttgaggctcgagcgaggcgga 213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 456 CGATGATGAGGAAGGGAGGGCCGCCATGAAGAGGAGGAGGTCGAGGCAGATGACCA 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 q 214
Cp 396 g 396
RESULT 10
LOCUS      RATNUCIA3      4204 bp      DNA      ROD      08-JAN-1991
DEFINITION Rat nucleolin gene.
ACCESSION M55020 M37081
NID        9205789
KEYWORDS   nucleolin.
SEGMENT    3 of 4
SOURCE     Rat DNA.
ORGANISM   Rattus norvegicus
            Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Rattus.
REFERENCE  1 (bases 1 to 4204)
AUTHORS    Bourbon,H.-M.M. and Analric.F.
TITLE      Nucleolin gene organization in rodents: Highly conserved sequences
            within three of the 13 introns
JOURNAL    Gene 88, 187-196 (1990)
MEDLINE    90269607
FEATURES    Location/Qualifiers
            1..4204

```

```

prim_transcript <1..>4202
              /organism="Rattus norvegicus"
intron        <1..4
              /note="nucleolin mRNA and introns"
exon          5..193
              /number=4
exon          294..380
              /number=5
intron        381..566
              /note="intron 4"
exon          567..708
              /number=6
intron        709..1189
              /note="intron 5"
exon          1190..1314
              /number=7
intron        1315..2067
              /note="intron 6"
exon          2068..2191
              /number=8
intron        2192..2699
              /note="intron 7"
exon          2700..2854
              /number=9
intron        2855..3312
              /note="intron 8"
exon          3313..3436
              /number=10
intron        3437..3626
              /note="intron 9"
exon          3627..3760
              /number=11
intron        3761..>4202
              /note="intron 10"
BASE COUNT   1307 a 695 c 915 g 1287 t
ORIGIN
Query Match      2.0%; Score 27; DB 91; Length 4204;
Best Local Similarity 73.7%; Pred. No. 3.50e-01;
Matches 42; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Db 118 gatgaagatgaaggaggatgaagatgaagatgaagatgaagatgaagatgaagatgag 174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 455 GATGATGAGGAAGGGAGGGCCGCCATGAAGAGGAGGTCGAGGCAGATGAG 399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 11
LOCUS      MUS45SRNA      4251 bp      DNA      ROD      22-SEP-1986
DEFINITION Mouse 4.5S rRNA gene.
ACCESSION M12658
NID        g191523
KEYWORDS   4.5S ribosomal RNA; repeat region.
SOURCE     Mouse DNA.
ORGANISM   Mus musculus
            Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
            Murinae; Mus.
REFERENCE  1 (bases 1 to 4251)
AUTHORS    Schoeniger,L.O. and Jelinek,W.R.
TITLE      4.5S rRNA is encoded by hundreds of tandemly linked genes, has a
            short half-life, and is hydrogen bonded in vivo to poly(A)
            terminated RNAs in the cytoplasm of cultured mouse cells
JOURNAL    Mol. Cell. Biol. 6, 1508-1519 (1986)
MEDLINE    87064431
COMMENT    Draft entry and sequence in computer readable form for [1] kindly
            provided by W.R.Jelinek, 28-JUL-1986.
            The 4.5S rRNA has a short half-life and appears to be associated
            with poly-A mRNAs in vivo. It is not a ribosomal RNA. There are
            approximately 850 copies of the 4.5S rRNA gene per haploid mouse
            erythrocyte cell genome. The DNA fragment presented here is a
            monomer unit of a tandem repeating array. The sequence contains an
            unusually high number of short tandem repeats (positions 75-119,

```







\*\*\*\*\*  
M P E L H  
(TM)  
\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MParch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:25:03 1997; Maspar time 8.34 Seconds  
Tabular output not generated.  
791.390 Million cell updates/sec

Title: >US-08-842-827-6  
Description: (1-311) from US08842827.pep  
Perfect score: 2326  
Sequence: 1 MQNKKYDKAIVPESKNGSP.....RKEILSPVDIIDRNNHNM 311

Scoring table: PAM 150  
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11  
Statistics: Mean 47.573; Variance 83.956; scale 0.567

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Description	Pred. No.
1	431	18.5	YX33_CAEEL	HYPOTHETICAL 39.0 KD
2	122	5.2	GU01_RAT	POSSIBLE GUSTATORY RE 1.43e-04
3	120	5.2	ATC5_YEAS	PROBABLE CALCIUM-TRAN 2.91e-04
4	118	5.1	YG1P_YEAS	HYPOTHETICAL 27.6 KD 5.91e-04
5	111	4.8	EDGL_MOUSE	PROBABLE G PROTEIN-CO 6.63e-03
6	111	4.8	H218_RAT	PROBABLE G PROTEIN-CO 6.63e-03
7	112	4.8	NUAM_CAEEL	NADH-UBIQUINONE OXIDO 4.72e-03
8	103	4.4	YBAR_BACSV	HYPOTHETICAL 46.4 KD 9.31e-02
9	100	4.3	VU25_HSV7J	U25 PROTEIN.
10	101	4.3	YIC3_YEAS	HYPOTHETICAL 37.4 KD 1.76e-01
11	99	4.3	YSCU_YERPS	YOP PROTEINS TRANSLOC 3.31e-01
12	101	4.3	LMPI_EBVC	LATENT MEMBRANE PROTE 1.76e-01
13	101	4.3	YAC1_LEGPN	POTASSIUM CHANNEL PRO 1.76e-01
14	97	4.2	SRG3_CAEEL	HYPOTHETICAL 23.7 KD 6.15e-01
15	97	4.2	LMPI_EBV	LATENT MEMBRANE PROTE 6.15e-01
16	97	4.2	LMPI_EBVC	LATENT MEMBRANE PROTE 6.15e-01
17	97	4.2	LMPI_EBVC	LATENT MEMBRANE PROTE 6.15e-01
18	98	4.2	NTRB_CANFA	SODIUM- AND CHLORIDE- 4.52e-01
19	97	4.2	YBA4_YEAS	HYPOTHETICAL 287.5 KD 6.15e-01
20	98	4.2	POLG_BVDVS	GENOME POLYPROTEIN.
21	96	4.1	LEP3_VIBCH	TYPE 4 PREPILIN-LIKE 8.35e-01
22	95	4.1	MIP_RANPI	LENS FIBER MAJOR INTR 1.13e-00

23	96	4.1	383	4	FLHB_SALTY	FLAGELLAR BIOSYNTHETI	8.35e-01
24	96	4.1	530	2	CIK6_RAT	POTASSIUM CHANNEL PRO	8.35e-01
25	93	4.0	184	11	YCF4_TOBAC	HYPOTHETICAL 21.5 KD	2.06e+00
26	93	4.0	207	7	NU6M_HANWI	NADH-UBIQUINONE OXIDO	2.06e+00
27	94	4.0	261	8	RNG6_HUMAN	CLASS II HISTOCOMPATI	1.53e+00
28	93	4.0	263	6	MIP_MOUSE	LENS FIBER MAJOR INTR	2.06e+00
29	92	4.0	326	11	YVW7_YEAS	HYPOTHETICAL 37.2 KD	2.77e+00
30	93	4.0	347	6	NU2M_BOVIN	NADH-UBIQUINONE OXIDO	2.06e+00
31	93	4.0	370	1	SH5B_MOUSE	5-HYDROXYTRYPTAMINE	5 2.06e+00
32	94	4.0	373	2	CARY_MOUSE	C-C CHEMOKINE RECEPTO	1.53e+00
33	92	4.0	492	4	GTR1_RABIT	GLUCOSE TRANSPORTER T	2.77e+00
34	92	4.0	492	4	GTR1_HUMAN	GLUCOSE TRANSPORTER T	2.77e+00
35	92	4.0	599	6	NTG1_RAT	SODIUM- AND CHLORIDE-	2.77e+00
36	92	4.0	599	6	NTG1_HUMAN	SODIUM- AND CHLORIDE-	2.77e+00
37	91	3.9	260	4	HA2Q_HUMAN	HLA CLASS II HISTOCOM	3.71e+00
38	91	3.9	261	6	MIP_RAT	LENS FIBER MAJOR INTR	3.71e+00
39	90	3.9	261	2	COX3_HUMAN	CYTOCHROME C OXIDASE	4.95e+00
40	91	3.9	275	7	POTB_ECOLI	SPERMIDINE/PUTRESCINE	3.71e+00
41	91	3.9	451	4	GRL1_PIG	GLUCOSE TRANSPORTER T	3.71e+00
42	91	3.9	492	4	GTR1_MOUSE	GLUCOSE TRANSPORTER T	3.71e+00
43	91	3.9	492	4	GTR1_BOVIN	GLUCOSE TRANSPORTER T	3.71e+00
44	91	3.9	492	4	GTR1_RAT	GLUCOSE TRANSPORTER T	3.71e+00
45	91	3.9	574	11	YHGE_ECOLI	HYPOTHETICAL 64.6 KD	3.71e+00

ALIGNMENTS

RESULT 1  
AC YX33\_CAEEL STANDARD; PRT; 341 AA.  
Q10022;  
DI 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DI 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 39.0 KD PROTEIN T28D9.3 IN CHROMOSOME II.  
GN T28D9.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: WEAK, TO YEAST D9719.9.  
DR WORMPEP; T28D9.3; CE02068.  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 71 91 POTENTIAL.  
FT TRANSMEM 122 142 POTENTIAL.  
FT TRANSMEM 223 243 POTENTIAL.  
FT TRANSMEM 257 277 POTENTIAL.  
SQ SEQUENCE 341 AA; 39028 MW; 89AE6E81 CRC32;

Query Match 18.5%; Score 431; DB 11; Length 341;  
Best Local Similarity 31.7%; Pred. No. 1.55e-66;  
Matches 83; Conservative 71; Mismatches 96; Indels 12; Gaps 10;

Db	28	islfflftaavtviptllgvsgqgffcdsddsiyeyrk-dtitavqlmlylnlnaa	86
Qy	39	LOLFCLFMAGLEPLFIETSTIKPVRHGFYCNDESIKYPLKRTGETINDAVLCAGVIVAIL	98
Db	87	tvlfeyvymqkvesninpvrnnhhlvfrlitygysqigfvmialniytkhv	146
Qy	99	AITGEFYR---I-YLKKSSTONPYVAALY-KQVGCFLFG-CAISQST-DI-AKVS	150
Db	147	vgrlrphfidvcklandctvtdgshrytdyctgppelvlearksfysghsavlscat	206
Qy	151	IGRLRPHFLSVCN-PDFSQINC-SEGVIYRCRGDSDSKVQEARKEFSGHAFSMYTML	208
Db	207	wealyiqarlgpvlrnriivpsotlcmfmgilglsfridcnkhshsdvlgvlgifila	266
Qy	209	YLVLYLQARFTWR-GARLLRPLQLTLLIMAFYGLSRVSDKHHPDVLGAFQAQALVA	267

[illegible]

```

FT TRANSMEM      35   59          1 (POTENTIAL).
FT DOMAIN        60   66          CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     67   95          2 (POTENTIAL).
FT DOMAIN       96   108         EXTRACELLULAR (POTENTIAL).
FT TRANSMEM    110  128         3 (POTENTIAL).
FT DOMAIN     129  147         CYTOPLASMIC (POTENTIAL).
FT TRANSMEM    148  173         4 (POTENTIAL).
FT DOMAIN     174  189         EXTRACELLULAR (POTENTIAL).
FT TRANSMEM    190  210         5 (POTENTIAL).
FT DOMAIN     211  233         CYTOPLASMIC (POTENTIAL).
FT TRANSMEM    234  255         6 (POTENTIAL).
FT DOMAIN     256  271         EXTRACELLULAR (POTENTIAL).
FT TRANSMEM    272  292         7 (POTENTIAL).
FT DOMAIN     293  352         CYTOPLASMIC (POTENTIAL).
FT CARBOHYD     19   19         POTENTIAL.
FT LIPID        305  305         PALMITATE (BY SIMILARITY).
SQ SEQUENCE     352 AA; 38734 MW; 2F532027 CRC32;

Query Match            4.88; Score 111; DB 4; Length 352;
Best Local Similarity 37.88; Pred.No. 6.63e-03;
Matches 17; Conservative 8; Mismatches 19; Indels 1; Gaps 1;

Db 191 vlcvtfsvllaivaalvalryivfvrrshadvagptlallktv 235
||| || | | : : ||| : : | : : | : | : | : |
QY 87 VLCAVGIVIAIATIGTEFRVIYL-KKRSIQNPYAALYKQV 130
||||| | | : : ||| : : | : : | : | : | : |

RESULT 7
ID NU4M_CAEEL STANDARD; PRT; 409 AA.
AC P24892;
DT 01-MAR-1992 (REL. 21, CREATED)
DD 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DI 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DN NADH + UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).
GN ND4.
OG CAENORHABDITIS ELEGANS.
OC MITOCHONDRION.
EO EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RX MEDLINE; 92201635.
RA OKIMOTO R.; MACFARLANE J.L., CLARY D.O., WOLSTENHOLME D.R.;
RL GENETICS 130:471-498(1992).
CC -! CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CR EMBL; X54252; G559502; -.
DR PIR; S26033; S26033.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SW SEQUENCE 409 AA; 47206 MW; DC040448 CRC32;

Query Match            4.88; Score 112; DB 7; Length 409;
Best Local Similarity 27.68; Pred.No. 4.72e-03;
Matches 21; Conservative 25; Mismatches 22; Indels 8; Gaps 7;

Db 18 flftvmfslflnn-fswggllfvldsyssillmvslfilgilvivise-knn-nlli- 72
| : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 202 FSNMYTMLVLVLYLQARTWRGALLRPLAQFL-IWMAFYT-GLSRVS DHKHHPSDVIAG 259
||| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 73 lse-ilvficiifip 87
::: || ||| : :

QY -260 FAQGALVACCICVFVFS 275
||| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
ID YBAR_BACSU STANDARD; PRT; 434 AA.
AC P55189;
DT 01-OCT-1996 (REL. 34, CREATED)
DD 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DI 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DN HYPOTHETICAL 46.4 KD PROTEIN IN RRNG-FEUC INTERGENIC REGION.
GX YEAR.
GO BACILLUS SUBTILIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
```



CC CC -!- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS CHARACTERIZED  
BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC CC -!- THE TAIL MAY BE IMPORTANT IN MODULATION OF CHANNEL ACTIVITY AND/OR  
TARGETING OF THE CHANNEL TO SPECIFIC SUBCELLULAR COMPARTMENTS.  
CC CC -!- THIS CHANNEL PROTEIN BELONGS TO THE DELAYED RECTIFIER CLASS.  
CC CC -!- SIMILARITY: BELONGS TO SHAKER POTASSIUM CHANNEL SUBFAMILY.

EMBL: X17622; G32033; -.  
DR DR PIR: S15057; S15057.  
DR DR MIR: J76257; -.  
DR KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;  
KW GLYCOPROTEIN; MULTIGENE FAMILY; PHOSPHORYLATION.  
FT FT TRANSMEM 175 193 SEGMENT S1.  
FF FT TRANSMEM 263 284 SEGMENT S2.  
FT FT TRANSMEM 296 316 SEGMENT S3.  
FT FT TRANSMEM 340 360 SEGMENT S4.  
FT FT TRANSMEM 376 396 SEGMENT S5.  
FT FT TRANSMEM 437 457 SEGMENT S6.  
FT FT MOD\_RES 511 511 PHOSPHORYLATION (BY CAPK)  
SQ SQ SEQUENCE 529 AA; 58728 MW; 356774CE CRC32;

Query Match 4.3%; Score 101; DB 2; Length 529;  
Best Local Similarity 20.6%; Pred. No. 1.76e+01;  
Matches 20; Conservative 27; Mismatches 46; Indels 4; Gaps 4;

Dd Db 428 mymptvggkivgs-lcaiaqvlialpvpiwvsnfnyfhreteeeqqgythvtcqpqa 486  
QY QY 74 KYPLKTGETINDAVLCavgviAI-LAI-ITGEFRYYLKRSSTIONPYVAALYKRGV 131  
Dd Db 487 polratdnlgikpfpeanier-rpsyltphrayae 522  
QY QY 132 CFLFGCAISQSQTIDIAKVSGRLRPHELSCVPNDFSQ 168

RESULT 14  
ID YACL\_LEGPN STANDARD; PRT: 208 AA.  
AC P37033:  
DT DT 01-JUN-1994 (REL. 29, CREATED)  
DD DD 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DE DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DS DS HYPOTHETICAL\_23\_7\_KD\_PROTEIN\_IN\_ACN\_5\_REGION.  
OS LEGIONEELLA PNEUMOPHILA.  
OC PRORARIOTPA; GRACILLICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
RN [1]  
RC STRAIN-PHILODELPHIA 1;  
RX MENGAUD J.M., HORWITZ M.A.;  
RZ J. BACTERIOL. 175:5656-5676(1993).  
DL EMBL: L22081; G348944; -.  
DR DR FIP: A48642; A48642.  
SW SW HYPOTHETICAL PROTEIN.  
XQ SEQUENCE 208 AA; 23714 MW; CEICOAB4 CRC32;

Query Match 4.2%; Score 97; DB 10; Length 208;  
Best Local Similarity 34.1%; Pred. No. 6.15e+01;  
Matches 14; Conservative 12; Mismatches 14; Indels 1; Gaps 1;

Dd Db 89 aihsykhkfyyyp-vahdklkpdoptenaaiatglvsalll 128  
QY QY 58 TKPPIHRGFCYNDESIRPKLTGETINDAVLCavgviTAIL 98

RESULT 15  
ID SRG3\_CAEEL STANDARD; PRT: 332 AA.  
AC P46572:  
DT DT 01-NOV-1995 (REL. 32, CREATED)  
DD DD 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DE DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DS DS SRG-3 PROTEIN.  
GN GN SRG-3 OR C18F10.6.

```

OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELONOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA LATREILLE P.;
RL SUBMITTED (MAY-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE SRG FAMILY OF C.ELEGANS RECEPTOR-LIKE
CC PROTEINS.
DR EMBL; U00049; G1216304; -.
KW TRANSMEMBRANE; MULTIGENE FAMILY.
FT TRANSMEM 23 43
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
SQ SEQUENCE 332 AA; 38886 MW; DBOC6C73 CRC32;

Query Match 4.2%; Score 97; DB 9; Length 332;
Best Local Similarity 29.6%; Pred. No. 6.15e-01;
Matches 32; Conservative 27; Mismatches 38; Indels 11; Gaps 10;

Db 146 livifla--pfliwvnlldsnfig-yvnggfgisyrvtwaslslmgftliitvlit 202
|: :||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 41 LFCLEFAGLPFLIITSTIKPYHRGFCNDS-SIKYPLK-T-GE-TINDAVLCVGVIA 96
|: :||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 203 mvtt-tvtfykmtnkrikaseralciaaalisvg-fille-aitgsf 247
|: :||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 97 ILAIITGEYRIYLLKKS-RSTIONPYVAALYKQVGCFLGCAISQSF 143
|: :||| |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: Tue Nov 4 10:25:51 1997  
Job time : 48 secs.



\*\*\*\*\*

MAJESTIC (TM)

\*\*\*\*\*

Release 2.1d John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run On: Thu Nov 6 12:23:32 1997; MasPar time 1029.95 Seconds  
Tabular output not generated. 1313.733 Million cell updates/sec

Title: >US-08-842-827-7  
Description: (1-1232) from US08842827.seq  
Perfect Score: 1232  
N.A. Sequence: 1 ACCATGACGCGAGGTGGT.....CAAAAAAAAAAAAAAAAA 1232  
Comp: TGGTACGTCGCTCACCCA.....GTTTTTTTTTTTTTTTTTTT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-new3  
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV  
9: ORG 10: NAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC  
17: VIR

Database: genbank99  
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7  
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2  
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3  
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10  
45: INV11 46: NAM1 47: NAM2 48: NAM3 49: VRT1 50: VRT2 51: VRT3  
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG  
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7  
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2  
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9  
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15  
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7  
92: ROD8 93: STR 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3  
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9  
105: VRL10  
Database: genbank-new3  
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV  
112: NAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2  
118: ROD 119: SYN 120: UNA 121: VRL  
Database: u-emb150\_99  
122: part1

Statistics: Mean 11.574; Variance 6.325; scale 1.830

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	127	10.3	1212	85	D84376	House mouse: Musculus	1.26e+66
2	119	9.7	2206	91	RNDRI42	R.norvegicus mRNA for	7.79e-61
3	106	8.6	1444	78	HSU73294	Human clone 23748 MRN	1.64e-51
4	59	4.8	1490	88	MUSHPIP	Mus musculus (clone H	2.95e-19
5	37	3.0	1622	40	DMU73822	Drosophila melanogast	8.42e-06
6	36	2.9	215	57	128278	Sequence 5 from paten	3.05e-05
7	33	2.7	215	57	128278	Sequence 5 from paten	1.33e-03
8	29	2.4	1104	80	HUMCOOTAA	Human CO-029.	1.58e-01
9	28	2.3	943	92	S42292	mannose-binding prote	4.94e-01
10	28	2.3	1584	52	XLHBOX8	Xenopus laevis XLHbox	4.94e-01
11	28	2.3	2770	91	RNPFKL	Rat PFK-L mRNA for 11	4.94e-01
12	28	2.3	38476	116	HUMU54B12	Human cosmid U54B12,	4.94e-01
13	28	2.3	38476	83	HUMU54B12	Human cosmid U54B12,	4.94e-01
14	28	2.3	41611	74	HSF77D12	Human DNA sequence fr	4.94e-01
15	28	2.3	117561	34	HS426N21	Human DNA sequence **	4.94e-01
16	28	2.3	242825	76	HSPEX	H.sapiens PEX gene.	4.94e-01
17	27	2.2	1108	46	BTIN6MR	B.taurus interleukin-	1.51e+00
18	27	2.2	1118	59	ATP22A	A.thaliana mRNA for p	1.51e+00
19	27	2.2	1160	91	RN17BHD1	R.norvegicus mRNA for	1.51e+00
20	27	2.2	1322	74	HSLGACAR	H.sapiens mRNA for no	1.51e+00
21	27	2.2	1717	90	MUSUHSK	Mus musculus ultra hi	1.51e+00
22	27	2.2	1951	68	STBINPP	S.tuberosum mRNA for	1.51e+00
23	27	2.2	3200	60	BNMYBIPRO	B.napus mRNA for myro	1.51e+00
24	27	2.2	36273	79	HUM7501	Homo sapiens chromoso	1.51e+00
25	27	2.2	44519	32	CSD1086	*** SEQUENCING IN PRO	1.51e+00
26	27	2.2	105576	70	HS106120B	Human DNA sequence fr	1.51e+00
27	27	2.2	124065	34	HS440021	Human DNA sequence **	1.51e+00
28	27	2.2	135048	6	HSU91325	Human chromosome 16p1	1.51e+00
29	27	2.2	135048	116	HSU91325	Human chromosome 16p1	1.51e+00
30	27	2.2	215285	116	HSU91322	Human chromosome 16p1	1.51e+00
31	27	2.2	215285	6	HSU91322	Human chromosome 16p1	1.51e+00
32	26	2.1	354	8	ORU89259	Oxytricha fallax 57kd	4.51e+00
33	26	2.1	1341	91	RNRPA2	R.norvegicus mRNA for	4.51e+00
34	26	2.1	15000	117	HSUNGENE	H.sapiens ung gene fo	4.51e+00
35	26	2.1	18648	77	HSU18671	Human Stat2 gene, com	4.51e+00
36	26	2.1	28230	73	HSCG1160	Human cosmid CG1160 f	4.51e+00
37	26	2.1	54336	83	HUMXPDG1	Homo sapiens ERCC2 ge	4.51e+00
38	26	2.1	90368	34	HS65B7	Human DNA sequence **	4.51e+00
39	26	2.1	129252	116	HSU91327	Human chromosome 16p1	4.51e+00
40	26	2.1	129252	6	HSU91327	Human chromosome 16p1	4.51e+00
41	26	2.1	167880	6	HSU91323	Human chromosome 16p1	4.51e+00
42	26	2.1	195158	34	HS473B4	Human DNA sequence **	4.51e+00
43	26	2.1	199167	122	HS431A14	Human DNA sequence **	4.51e+00
44	26	2.1	199167	109	HS431A14	Human DNA sequence **	4.51e+00
45	26	2.1	220633	6	HSU91321	Human chromosome 16p1	4.51e+00

ALIGNMENTS

RESULT	1	D84376	1212 bp	mRNA	ROD	03-OCT-1996
LOCUS		House mouse; Musculus domesticus kidney mRNA for Phosphatidic acid phosphatase, complete cds.				
DEFINITION		D84376				
ACCESSION		gl487872				
NID		Phosphatidic acid phosphatase; 35-kDa phosphatidic acid phosphatase.				
KEYWORDS		Mus musculus kidney cDNA to mRNA.				
SOURCE		Mus musculus				
ORGANISM		Eukaryotae; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus				
REFERENCE		1 (bases 1 to 1212)				
AUTHORS		Kanoh, H.				
TITLE		Direct Submission				
JOURNAL		Submitted (10-APR-1996) to the DDBJ/EMBL/GenBank databases. Hideo Kanoh, Sapporo Medical University School of Medicine, Department of Biochemistry, West-17, South-1, Sapporo, Hokkaido 060, Japan (E-mail: kanoh@serpent.cc.sapmed.ac.jp, Tel:011-611-2111(ex.2290), Fax:011-612-5861)				

```

REFERENCE 2 (bases 1 to 1212)
AUTHORS Kai, M., Wada, I., Imai, S., Sakane, F. and Kanoh, H.
TITLE Identification and cDNA cloning of 35-kDa phosphatidic acid
phosphatase (type 2) bound to plasma membranes. Polymerase chain
reaction amplification of mouse H202-inducible hic53 clone yielded
the cDNA encoding phosphatidic acid phosphatase
J. Biol. Chem. 271 (31), 18931-18938 (1996)
JOURNAL 96324980
MEDLINE
FEATURES Location/Qualifiers
SOURCE 1..1212
/organism="Mus musculus"
/tissue_type="kidney"
287..1138
/evidence=experimental
/codon_start=1
/product="Phosphatidic acid phosphatase"
/translation="MEDKRLPYVALDVICVLLAGLFAILTSRHPFQRIQPCNDSDS
IKPIYEDIPALGGVIFPCIIIVMSIGSLSYVFNVLHNSFVGNPIATIIYKAV
GAEFLGVSASQSLDIARKYIGSLRPHFLAICNDWSKINCSDGYIEDIYICOGNEEKY
KEGLSFYSGHSFSFMYCMLFVALYQLMRGLRPLMLQFLAFIAYSIYVGLSRV
SDYKHNSDVTVCLIOGRAMAILVALYVSDFKDTSHYKKEKEDPHTTLHETASSRN
YSTNHP"
BASE COUNT 277 a 316 c 313 g 306 t
ORIGIN
Query Match 10.3%; Score 127; DB 85; Length 1212;
Best Local Similarity 59.8%; Pred. No. 1.26e-66;
Matches 434; Conservative 0; Mismatches 289; Indels 3; Gaps 2;

Db 313 cytgccctcgtatgtgattgctgtgctggtgattgctcttttgcattcttacttc 372
QY 24 CGTGTGCTCGACGTGTGTCTACTGTGCGCTCCCGTCCGCTATCCTGACGCT 83
Db 373 aaggaatcccccttcagcaggaataattctgaatgatgatccatcaagtagacctta 432
QY 84 GGTGAACCCCGGTACAAAGCGAGATTTTACTCGGGGGATGATCCATCCGGTACCCCTA 143
Db 433 caagggaagacaccattactgctctattagggtggaatgactcattctctgattat 492
QY 144 CGGTCCAGATACCATACCCACGGGTCTATGCTGGGGTCCACATACAGCCACCGTCAT 203
Db 493 cgtatgagtagtggagaactctgtctgttacttacttaactgctctgcatcgaattcctt 552
QY 204 CATTGCTCGCCGGGGAAGC-CTACCTGGTG--TACACAGACCGGCTCTATTCTCGCTC 260
Db 553 tgtcggcaatccctacatagccaccattttcaaaagccgtcgagcctttttgttcggagt 612
QY 261 GGACTTCAACAACATACGTGGCTGCTGTATACAAAGTGTCTGGGACCTTCTCTTTGGGGC 320
Db 613 ctacagtagtcagctcctgactgacatcgtaagtatactataggcagtttgcggcgca 672
QY 321 TGCCGTGAGCCAGTCTCTGACACCTTGGCCAAAGTACATGATTGGGCTCTGAAGCCCAA 380
Db 673 attcttgctatctgttaaccacagactgggtcaaaaatacaactgcagtgatggctattga 732
QY 381 CTTCCTAGCCGTCTGCGACCCGACCTGGAGCGGGTCAACTGCTCGGCTATGTGCAGCT 440
Db 733 ggaactacatgtcaaggggaatgaagaaagtaaggaagggcaggttgtttctactc 792
QY 441 GGAGNAGGTGTGCGAGGGAACCCCTGCTGATGTACCGAGGCCAGGTGTCTTTCTACTC 500
Db 793 gggacactcttcattctctctgtactgcagctgctttgttcgcaactttatcttcaagcag 852
QY 501 GGGACACTCTTCTCTTTGGGATGACTGCATGTGTCTTGTGGCGCTGTATGTGCAGGCAG 560
Db 853 gatgaagggagactggcagaagactcttagaccatgctcgaatttgggctcattgtttt 912
QY 561 ACTCTGTGGGAAGTGGCAGCGGTGCTGCGACCCACAGTCCAGTTCTTCTCGGTGGCGCT 620
Db 913 ttccatatatgtgggctcttctcgagtgtctgactacaaacaccactggagtcagtcac 972
QY 621 TGCCCTCTACGTGGGCTACACCCGCGTGTCTGATTACAAACACCACCTGGAGCGATGCTCT 680

```

```

Db 973 agtgggactcattcaggagcgtcgtatggctatatactggttctgtgtatgtatccgatt 1032
QY 691 TGTGGCTCTCTCGAGGGGCGACTGCTGGCTGCCCTACTGCTGCTACATCTCAGACTT 740
Db 1033 cttcaa 1038
QY 741 CTTCAA 746

RESULT 2
LOCUS RNDRI42 2206 bp RNA ROD
DEFINITION R.norvegicus mRNA for ER transmembrane protein. 22-NOV-1996
ACCESSION Y07783
NID g1684744
KEYWORDS Dri 42 gene; ER-transmembrane protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2206)
AUTHORS Barila, D., Plateroti, M., Nobili, F., Muda, A. O., Xie, Y., Morimoto, T.
and Perozzi, G.
TITLE The Dri 42 gene, whose expression is up-regulated during epithelial
differentiation, encodes a novel endoplasmic reticulum resident
transmembrane protein
J. Biol. Chem. 271 (47), 29928-29936 (1996)
REFERENCE 2 (bases 1 to 2206)
AUTHORS Perozzi, G.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1996) G. Perozzi, Istituto Nazionale Della
Nutrizione, Unit of Experimental Nutrition, Via Ardeatina 546,
00178 Roma, ITALY
FEATURES Location/Qualifiers
source 1..2206
/organism="Rattus norvegicus"
/strain="Wistar"
/cell_type="epithelial"
/dev_stage="adult"
/tissue_type="small intestine"
410..1348
/gene="Dri 42"
/codon_start=1
/product="ER transmembrane protein"
/db_xref="PID:g1684745"
/translation="MQSKYDKAIVPSKNGSGPALNNPRKGGSKRVLLICLDLFLCL
FMALPFLIIESTIKPIRGGFYCNDESIRYPLKVSSETINDAVLCVAGVILVAILRIIT
GSEFYIYIKERKSTIONPYAALYKQVCFLLGCAISQSFSDIAKYSIGLRPHFL
SYCDPDFSQINCEGYTONRCEDSKVQARKSFSGHAFSMFTMLYLYLYLAR
FTWGRALLRPLLOFTLLMFAFTGLSRVSDYKHPSDVLGAFAGALVACCVIFEVFV
DLFKYKTTLSLPAPAIRREITLSPVDMDRSNHHNMV"
BASE COUNT 503 a 634 c 583 g 486 t
ORIGIN
Query Match 9.7%; Score 119; DB 91; Length 2206;
Best Local Similarity 62.8%; Pred. No. 7.79e-61;
Matches 296; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

Db 776 tatgtggcgcctctctataagcaagtgggagctctctcttcggcgttccttcagcagcag 835
QY 274 TACGTGGCTCTGTATACAAAGTGTCTGGGACCTTCTCTGTTGGGGCTGCGGTGAGCCAG 333
Db 836 tccttcacagacatcgcgaagtgctccattggggcgcctcagggcctcacttcctcagcgtc 895
QY 334 TCTCTGACACACCTGGCCCAAGTACATGATTGGGCGCTCTGAAGCCCAACTTCTTAGCGGTC 393
Db 896 tgtgacctgatttcagtcagatcaattgctccgagggtcattcagaaactcacaggtgc 955
QY 394 TGCCACCCGACTGGAGCCGGGTCAACTGCTCGGTCTATGTGCAGCTGGAGAGGTGTGC 453

```

```

Db 956 cgaggagacagcaaaatcacagagccaggaatctctcttcgagccacgctcc 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 454 AGGGAAACCTGCTGATGTCACCGAGCCAGGTGCTTCTACTCGGACACTCTTCC 513

Db 1016 ttctcatgttcaaatgctgatatctggtgtttatctacagcccgcttcaactggcgt 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 514 TTTGGGATGTAAGTGTGTTGGCGTGTATGTGAGGACGACACTCTTGTGAAG 573

Db 1076 gggggccgctctccgccccctctgagttcaacttggctcatgagggcctctcacg 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 574 TGGGACGGCTGTCGACCCACAGTCCAGTTCCTCTGTTGGCTTGGCCCTACGTG 633

Db 1136 ggtattgcagtgatgactactactacaaacacactctcagcgtatgctggcagattgoc 1195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 634 GGTACACCGCGGTGCTGATTACAAACACCACTGGAGCGATGTCCTTGTGGCCCTG 693

Db 1196 caaggagcttggtggcctgctgcagtagttcttctggtgctgacctctcaa 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 694 CAGGGGCACTGGTGCCTCCCTACTGCTGCTACATCTCAGACTTCTTCAA 746

RESULT 3
LOCUS HSU79294 1444 bp mRNA PRI 30-JAN-1997
DEFINITION Human clone 23748 mRNA, complete cds.
ACCESSION U79294
NID 91710275
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
  1 (bases 1 to 1444)
  Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
  Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Andersson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.
TITLE A 'double adaptor' method for improved shotgun library construction
JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)
MEDLINE 96207227
REFERENCE
  2 (bases 1 to 1444)
  Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
  Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
  Large Scale Concatenation cDNA Sequencing
  Unpublished
JOURNAL
REFERENCE
  3 (bases 1 to 1444)
  Yu,W. and Gibbs,R.A.
  Direct Submission
  Unpublished
TITLE
  Submitted (22-NOV-1996) Molecular and Human Genetics, Baylor
  College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
JOURNAL
FEATURES
  Location/Qualifiers
  source
    1..1444
    /organism="Homo sapiens"
    /clone="23748"
    /sex="female"
    /dev_stage="infant"
    /tissue_type="brain"
    /clone_lib="Soares library INIB from IMAGE consortium"
    375..1076
    /codon_start=1
    /product="unknown"
    /db_xref="PID:g1710276"
    /translation="MONTYKDKAIVPEKNGSPALNNPRSGSKRVLLICLDLFCI
    EVAGLPFLIETSIKPHRGFYCNDSIKYPIKGTGETINDAVLCAGVIVAILIIT
    GFYRIYIKRSRTIQNPVVALYKQVCFLLGCAISQSFDDIAKVSIGRLRPHFLS
    TTRGARCSPGSCSP"
  CDS
    327 a 427 c 373 g 317 t
    Query Match 8.6%; Score 106; DB 78; Length 1444;
    Best Local Similarity 61.6%; Pred. No. 1.64e-51;
    Matches 295; Conservative 0; Mismatches 183; Indels 1; Gaps 1;

Db 732 aaccctacgtggcagcaactctataagcaagtgaggctgctctctcttggcgtgcatc 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 AACAACTACGTGCTGCTATACAAAGGTGCTGGGACCTTCCCTGTTTGGGCTGCGGTG 327

```

```

Db 792 agccagtctttcacagacattgccaaagtgtccatagggcgctggtctcactctcttg 851
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 AGCCAGTCTCTGACAGACTGTCGCAAGTACATGATTGGGCGTCTGAAGCCCAACTTCTTA 387

Db 852 agtgtgcgaacctgatttcagccagcatcaactgctctgaaaggctacatcagaactac 911
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 GCGCTGTCGACCCGACTGGAGCGGCTCAACTGCTCGGTCTATGTATGTCAGCTGGAGAAG 447

Db 912 agatgcagaggatgacagcaaaatccaggaagccaggaagtcctctctcttgccat 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 GTGTGAGGGGAAACCTCTGCTGATGTACCGAGGCCAGGTGCTTCTTCTACTCGGACAC 507

Db 972 gctctcttccatgtacactatgctgtatttgggtgtatatacctgcagggccgcttcaact 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 508 TCTTCTTTGGGATGTACTGTCATGTGTGTCTGTCTGCGGTGTATGTGACGACGACTCTGT 567

Db 1032 tggcagagagcccg-ctgctccggccctcctcgtcagttccacttgatcatgagccttc 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 568 TGAAGTGGGCACGGCTGCTGGACCCACACATCCAGTCTCTCTGTGGGCTTTGCCCTTC 627

Db 1091 tacacgggactgtctgcgtatcacagcaccacagcaccatcccagtgatgttcttggcagga 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 TAGCTGGGTACACCCGCGTGTCTGATTACAAACACCACTGGAGCGATGTCCTTGTGGC 687

Db 1151 ttgtctcaaggagccctggtggcctgctgcatagtttcttctgctgctgacctctcaa 1209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 688 CTCCTCAGGGGGCACTGGTGGCTGCCCTCTACTGTCTGTACATCTCAGACTTCTTCAA 746

RESULT 4
LOCUS MUSHPIP 1490 bp mRNA ROD 11-JAN-1996
DEFINITION Mus musculus (clone HIC-53) hydrogen peroxide-inducible protein
ACCESSION L43371
NID g1161099
KEYWORDS hydrogen peroxide; hydrogen peroxide-inducible protein.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE
  1 (bases 1 to 1490)
  Egawa,K., Yoshiwara,M., Shibamura,M. and Nose,K.
  Isolation of a novel ras-recision gene that is induced by hydrogen
  peroxide from a mouse osteoblastic cell line, MC3T3-E1
  FEBS Lett. 372 (1), 74-77 (1995)
JOURNAL 96032549
MEDLINE
FEATURES
  Location/Qualifiers
  source
    1..1490
    /organism="Mus musculus"
    /clone="HIC-53"
    /cell_line="MC3T3-E1"
    /cell_type="osteoblast"
    /dev_stage="embryo"
    /tissue_type="calvariagag"
    <1..978
    /codon_start=1
    /product="hydrogen peroxide-inducible protein"
    /db_xref="PID:g1161100"
    /translation="ARASLELARVGEGTGRRPIGRDEPALREQAQIAASTVVK
    ESRKCSVRRPIQTGPQSRQSKRLAVAGRHPEGGALGIGLDRGLPLPLAP
    GDDTIQPMFDKRLPVLVDVICLLAGLPAILTSRHTFQRIKPYFCNDSDSIKPY
    KEDTIYALGGIVIPFCIVMSIGESLSVYENVHNSGVNPIYATIKYKAVGAFLE
    GVSASQSLDIKVTIGSLRPLAICNDPMKINGSDGVIYDIYICQNEKVEKRL
    SPSGTILFYLVLHVAVCRLSSRLRLRLRPLMLOFGLIATSIYVGLSSV"
  CDS
    979..1490
    362 a 367 c 368 g 393 t
    Query Match 4.8%; Score 59; DB 88; Length 1490;
    Best Local Similarity 57.9%; Pred. No. 2.95e-19;

```



QY	1165	AAATGTAATGTATGATGGTGTCTTTTACTAATAAGG-GCACCTGTTCACAAAAAAA	1233
Db	1097	aaaaaaaa 1104 	
QY	1224	AAAAAAAA 1231	
RESULT	9		
LOCUS	S42292	943 bp mRNA	ROD 16-NOV-1992
DEFINITION	mannose-binding protein A precursor=acute phase reactant [mice, Inbred CBA/J, acute phase liver library, pTZ 19 vector, mRNA, 943 nt].		
ACCESSION	S42292		
NID	NID		
KEYWORDS	g1679939		
SOURCE	Mus sp.	Inbred CBA/J pTZ 19 vector acute phase liver library.	
ORGANISM	Mus sp.		
Eukaryotes;	mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1	(bases 1 to 943)	
AUTHORS	Sastry,K., Zahedi,K., Lelias,J.M., Whitehead,A.S. and Ezekowitz,R.A.		
TITLE	Molecular characterization of the mouse mannose-binding proteins. The mannose-binding protein A but not C is an acute phase reactant		
JOURNAL	J. Immunol.	147 (2), 692-697 (1991)	
MEDLINE	91302823		
REMARK	Genbank staff at the National Library of Medicine created this entry [NCBI gibbsq 4292] from the original journal article. This sequence comes from Figure 2.		
FEATURES	Location/Qualifiers		
source	1..943	/organism="Mus sp."	
CDS	121..840	/note="Description: mannose-binding protein A precursor, MBP-A precursor; acute phase reactant; This sequence comes from Figure 2"	
		/codon_start=1	
		/product="mannose-binding protein A precursor"	
		/db_xref="PID:g233016"	
		/translation="MLLLPLPVLCVSVSSSGSTCEDTLKTCVSIACGRDGRGFGKGEPSGLRGQLGPGPKLGPCKLGPPGSVPSPGPKGKGHDGNRAETEERLANNAEIRILXSKLOLTNLHAFSGKRSKKLFVTNHEKPMYSKVSLCTEQGTVAIPRNABENKAIQVAITGIAFLGITDEARTEQGFMIVIGRLIYSNKKRDEPNHHGGSEDCVIIID NGLMDISCOASFKACEFPA"	
BASE COUNT	270 a	217 c	207 t
ORIGIN			
Query Match	2.3%;	Score 28;	DB 92; Length 943;
Best Local Similarity	85.0%;	Pred. No. 4.94e-01;	
Matches	34;	Conservative	0; Mismatches 6; Indels 0; Gaps 0;
Db	901	aaatcagtcagtcttctctcaaaaaaaa	940
		c	
QY	1193	AAATAGGCCTGTTTTCACAAAAAAA	1232
RESULT	10		
LOCUS	XLHXO8	1584 bp DNA	VRT 11-AUG-1994
DEFINITION	Xenopus laevis	xlhx8 gene for homeodomain protein.	
ACCESSION	X15849		
NID	9530782		
KEYWORDS	developmental regulation; DNA-binding protein; homeobox; transcription factor; xlhx8 gene.		
SOURCE	African clawed frog.		
ORGANISM	Xenopus laevis		
Eukaryotes;	mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.		
REFERENCE	1	(bases 1 to 1584)	
AUTHORS	Wright,C.V.E., Schnegelsberg,P. and De Robertis,E.M.		
TITLE	xlhx8 : a novel Xenopus homeo protein restricted to a narrow band of endoderm		

```

JOURNAL Development 104, 787-794 (1988)
REMARK revised by [2]
REFERENCE 2 (bases 1 to 1584)
AUTHORS Gama, L.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-1994) Laura Gama, 615-343-8258 (lab) 615-343-4539
COMMENT (fax) e-mail: wrightc@ctvax.vanderbilt.edu
FEATURES Data kindly reviewed (22-JUN-1990) by De Roberts E.M.
Location/Qualifiers
Source 1..1584
/organism="Xenopus laevis"
/clone_lib="Charon 4A"
70..885
/notes="xlhbox 8 protein"
/codon_start=1
/db_xref="pid:g530783"
/translation="MNADQYEQAPIKEPCAFORSQGDYNPSPACILYMGROQQA
AYSNPLVALDGGSPDISPYEPPSEEDFYPLHHHHHHHHHHPGIPHPHQMPPF
DNDSGGLERNTLLPFPWMSKTSKSHWGQWDGSTMEEQENKRTAYTRAQLL
ELEKEFLNFYIISRPRLVELAMNLNTERHIKINFONRRMKKEEDKXRGSDPEQ
DSVSSADVLKDEPQGLGSKQKTGDLVLSPLTSSQPNQVPSIGSLRQAEK"
BASE COUNT 527 a 342 c 303 g 412 t
ORIGIN
Query Match 2.3%; Score 28; DB 52; Length 1584;
Best Local Similarity 79.2%; Pred. No. 4.94e-01;
Matches 36; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 1526 tttttattataaagcacattgttgaataaaataaaataaaataaa 1573
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1185 TTTTGTAGTAATAGGCACCTGTTTCACAAAAAATAAATAAATAA 1232

RESULT 11
LOCUS RNPFKL 2770 bp RNA ROD 16-DEC-1991
DEFINITION Rat PFK-L mRNA for Liver Phosphofructokinase.
ACCESSION X58686
NID 956886
KEYWORDS phosphofructokinase.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Rattus.
REFERENCE 1 (bases 1 to 2770)
AUTHORS Hotta,K.
TITLE Direct Submission
JOURNAL Submitted (03-APR-1991) K. Hotta, Second Dept of Internal Medicine,
Osaka University Medical School, 1-1-50 Fukushima, Fukushima-ku,
Osaka 553, JAPAN
REFERENCE 2 (bases 1 to 2770)
AUTHORS Hotta,K., Nakajima,H., Yamaaki,T., Hamaguchi,T., Kuwajima,M.,
Noguchi,T., Tanaka,T., Kono,N. and Tarui,S.
TITLE Rat-liver-type phosphofructokinase mRNA. Structure, tissue
distribution and regulation
JOURNAL Eur. J. Biochem. 202 (2), 293-298 (1991)
MEDLINE 92104147
FEATURES Location/Qualifiers
Source 1..2770
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/tissue_type="liver"
/clone_lib="lambda-gt10"
/clone="lambda-PFKL1-PFKL8"
1..2740
/evidence=experimental
/gene="PFK-L"
62..70
/gene="PFK-L"
67..2409
/gene="PFK-L"
/EC_number="2.7.1.11"

```

---

```

/codon_start=1
/product="6-phosphofructokinase"
/db_xref="pid:g56887"
/db_xref="SWISS-PROT:P30835"
/translation="NATVDEKLRMSGAGRAIGVLTSGDAQGNAAVRAVTRMGIVV
GAKVFIYEGYEGVLVEGGENIKPANLWSNIIOLGTTIIGSARCAFTTREGRLAA
YNLQHGITNLGVIGDGLSLGCANIFRNEMGSLLELVKSGKISESTANVAHLSTAG
LVGSDINDFCGDTMCIQDSALHRIEMVIDAITTTAQSHQRTVLEVMRGHGYLALV
SALASADNLFIPEAPPEDGEGNEMFCERIGETSRGSLNIIIIAGADIDRHGPIISS
SYVLDLVQRIGDFRTVTVLGRVQGGTSPAFDRVLSKRMGMEAVMALDEIDPTTAC
VYLSGNSQSVRLPLMECVQVTDVQKAMDEKREDEAIOLGRSFENNNKIYKLLAHQK
VSKESNFSLAIIYNGAPAGMNAARSAVETGISEGTYVYVVDGFEGLAKGVQVEV
GHDVAGNLGRGSGMLGTRTLPKPHLEAIVENLRTYNIHALLVIGGFEAYEGVLQV
EARGRYELCIVMCVIPATISNNVPGTDFSLGSDTAVNAAMESCDRIKOSASCTKRVV
LRNEKCHEYTFEFLNLYSSEGCVDFCRNVLGHLOQGGAPTDPDRNYGTRKLGVA
MLMWSKLRDVRKRVFANAFDSVCYGLKRVVAFSSVTEKKETDFEHRMPRQDM
WMLRLMLKMLAHYRISMADYVSGELEHVTRTLSDIKGP"
2722..2727
/polyA_signal
/gene="PFK-L"
2740
/polyA_site
/gene="PFK-L"
BASE COUNT 609 a 762 c 864 g 535 t
ORIGIN
Query Match 2.3%; Score 28; DB 91; Length 2770;
Best Local Similarity 79.2%; Pred. No. 4.94e-01;
Matches 36; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 2713 tttttagaataaagcacctgttttagaataaaataaaataaaataaa 2760
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1185 TTTTGTAGTAATAGGCACCTGTTTCACAAAAAATAAATAAATAA 1232

RESULT 12
LOCUS HUMU54B12 38476 bp DNA PRI 12-FEB-1997
DEFINITION Human cosmid U54B12, complete sequence.
ACCESSION U73024
NID G1613887
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 38476)
AUTHORS Fulton,B. and Rohlfing,T.
TITLE The sequence of H. sapiens cosmid U54B12
JOURNAL Unpublished (1996)
AUTHORS 2 (bases 1 to 38476)
TITLE Direct Submission
JOURNAL Submitted (01-OCT-1996)
AUTHORS 3 (bases 1 to 38476)
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1996)
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
e-mail: sapiens@watson.wustl.edu

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was











\*\*\*\*\*



\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by Intelligenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 12:43:31 1997; Maspar time 132.50 Seconds

Tabular output not generated. 970.443 Million cell updates/sec

Title: >US-08-842-827-7

Description: (1-1232) from US08842827.seq

Perfect Score: 1232

N.A. Sequence: 1 ACCATGCGACGAGTGCGT.....CAAAAAAAAAAAAAAAAA 1232

Comp: TGGTACGTCGCTCCACCCA.....GTTTTTTTTTTTTTTTTT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 142080 seqs, 52183452 bases x 2

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: n-geneseq28

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29

Statistics: Mean 9.327; Variance 6.638; scale 1.405

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	83	6.7	1047	2	Q10572	1.19e-29
2	75	6.1	1047	2	Q10572	2.59e-25
3	70	5.7	74	20	T23564	1.32e-22
4	42	3.4	91	9	Q51746	3.66e-08
5	42	3.4	91	9	Q51746	3.66e-08
6	42	3.4	204	1	N81154	3.41e-07
7	40	3.2	204	1	N81154	2.71e-05
8	36	2.9	114	12	Q70467	2.29e-04
9	34	2.8	114	12	Q70465	2.29e-04
10	34	2.8	114	12	Q70466	2.29e-04
11	34	2.8	114	12	Q70469	2.29e-04
12	34	2.8	114	12	Q70468	2.29e-04
13	34	2.8	114	12	Q70468	2.29e-04
14	34	2.8	114	12	Q70469	2.29e-04
15	33	2.7	81	21	T13611	6.57e-04

C	16	33	2.7	114	12	Q70467	Generic DNA sequence	6.57e-04
C	17	32	2.6	114	12	Q70472	Generic DNA sequence	1.86e-03
C	18	32	2.6	114	12	Q70465	Generic DNA sequence	1.86e-03
C	19	32	2.6	114	12	Q70466	Generic DNA sequence	1.86e-03
C	20	31	2.5	67	24	T14322	Primer used in the la	5.23e-03
C	21	30	2.4	74	21	T13613	DC43 TSAR library gen	1.45e-02
C	22	30	2.4	114	12	Q70470	Generic DNA sequence	1.45e-02
C	23	30	2.4	114	12	Q70471	Generic DNA sequence	1.45e-02
C	24	30	2.4	114	12	Q70470	Generic DNA sequence	1.45e-02
C	25	29	2.4	114	12	Q70472	Generic DNA sequence	3.97e-02
C	26	29	2.4	1104	3	Q22438	CO-029 tumour associa	3.97e-02
C	27	28	2.3	114	12	Q70473	Generic DNA sequence	1.07e-01
C	28	27	2.2	114	12	Q70471	Generic DNA sequence	2.86e-01
C	29	27	2.2	775	25	T31973	Human RNA polymerase	2.86e-01
C	30	26	2.1	114	12	Q70473	Generic DNA sequence	7.51e-01
C	31	26	2.1	198	7	Q42784	Ligand-induced gene,	7.51e-01
C	32	26	2.1	1031	18	O99786	Plant SAR gene pDPA2.	7.51e-01
C	33	26	2.1	1705	11	O62531	Allinase cDNA derived	7.51e-01
C	34	26	2.1	2371	9	O50643	Human Ews gene clone	7.51e-01
C	35	26	2.1	3547	2	N60846	Plasmid sequence enco	7.51e-01
C	36	26	2.1	3871	2	N71302	HSV-1 gB and surround	7.51e-01
C	37	25	2.0	70	21	T17035	Human bax gene promot	1.94e+00
C	38	25	2.0	972	29	T62760	Human mitochondrial D	1.94e+00
C	39	25	2.0	1139	1	N80299	Interleukin 6.	1.94e+00
C	40	25	2.0	1788	12	O73384	IT10C3, transporter p	1.94e+00
C	41	25	2.0	3311	17	T10562	Alpha-amylose gene.	1.94e+00
C	42	25	2.0	4195	29	T68321	Human multidrug resis	1.94e+00
C	43	25	2.0	8174	29	T61677	Human alpha(1,2)-fuco	1.94e+00
C	44	25	2.0	8174	9	Q56908	DNA encoding a glycos	1.94e+00
C	45	25	2.0	53577	28	T18551	Human polycystic kidn	1.94e+00

## ALIGNMENTS

RESULT 1  
 ID Q10572 standard; DNA; 1047 BP.

DT 09-APR-1991 (first entry)  
 DE Human Natriuretic Peptide Receptor B.  
 KW NPB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;  
 KW hyperaldosteronism; glaucoma; guanyl cyclase.  
 OS Homo sapiens.

Key: Location/Qualifiers

FT Peptide 1..22  
 FT /label= signal sequence  
 FT Protein 12  
 FT /label= mature NPB  
 FT Domain 23..455  
 FT /label= extracellular domain  
 FT /note= "binds natriuretic peptides A, B and C]"  
 FT Domain 456..456  
 FT /label= transmembrane domain  
 FT Domain 479..1047  
 FT /label= cytoplasmic domain  
 FT /note= "GC and protein kinase activity"  
 FT Modified -site 24..26  
 FT /label= N-glycos -site  
 FT Modified -site 35..37  
 FT /label= N-glycos -site  
 FT Modified -site 161..163  
 FT /label= N-glycos -site  
 FT Modified -site 195..197  
 FT /label= N-glycos -site  
 FT Modified -site 244..246  
 FT /label= N-glycos -site  
 FT Modified -site 277..279  
 FT /label= N-glycos -site  
 FT Modified -site 349..351  
 FT /label= N-glycos -site  
 FT Modified -site 600..602  
 FT /label= N-glycos -site  
 FT WO9100292-A.  
 PD 10-JAN-1991.



Matches	79; Conservative	258; Mismatches	575; Indels	10; Gaps	10;
Db	40	wawrvgnavanangranvndrvnrvsnngagacsnynannasvdknkyhdndnnngn	99		
Cp	1190	TAAAAACCATATATACATTACATTTTACAAAACAGCAATATCTGATCTCGTGCCCTT	1131		
Db	100	gcvyneasvarnashrnnntacavagsnsakndhyrtvrtgnsaunkngnvvtbhgh	159		
Cp	1130	CCTTAACCCATAAAAGAAGGGATTTTGGGGACCGACGGAACAGAGTTCCCCCTCCAA	1071		
Db	160	nnwtaraannyndardtdrhnbtongvnnannngsnsvnhvayarnvggnnnathnnra	219		
Cp	1070	ATGCTGAGGGCTACCCAGGCATCTCCAGACTCTGTGTCAGTGCAGGCGAGGGGCA-GC	1012		
Db	220	narnvncggnnmhnnnnnnaarntntgdyvnyndvngnsragnratrgnwdnrt	279		
Cp	1011	GGAAACGCTCACTGCTCCCATCAGCCAGCGGTTCTCTGGAGCCGCTCCAGAACCTCA-CT	953		
Db	280	rnnanarannntvntvyrnnnnnynnnnnnrnnrardnbgvngngsnmnnnagcnyd	339		
Cp	952	GCCTAACCAAGCTGAGGACCACTGGTGGGCATCAGCTGGACTCACAGCAGTCCTCT	893		
Db	340	gnnyanvnnntnongttrndgnrvnkmagryghvtnvnmkndrntdnvnnwangd	399		
Cp	892	GCCTGGCGGGGTCGGGC-CTCAGGAAGAAGATGCGGGTATCCCATAGTG-GTGTGG	835		
Db	400	ndsgdnnaahysganknwvtrgnnwvxnnsdnncdndndndscdktnstnanv	459		
Cp	834	TCAGCCTCGCCCGGTCACAGCTCAGTCAGCAGCGCTGGGCTCTCCGTCCAGCTCCTCTC	775		
Db	460	angtgnntnmgyssnnnrkmmnknaasmwrrvnnnnngnsryhkgsgsrntsn	519		
Cp	774	CTTCAGACAGTCTGTGGGGTCGGGCTTTGAAGAAGTCTGAGA-TGTAGCAGACAGTGA	716		
Db	520	rgssygnmtahgkynnantghknvnavankhvknkrrnttrvnnnkhndvnnhn	579		
Cp	715	GGGACGCCACCACTGCCCTCTGCAGGAGGCCAACAGGACATCGCTCCAGTGGTGT-TG	657		
Db	580	trngacndnnncvtnvncrgsnnndnnndnnndnmrysnnndvkgmannhnsnn	639		
Cp	656	TAACTAGACACGCGGGTGT-AGCCCCAGTAGAGGGCAAGGCCACACGAAGAATCGAC	598		
Db	640	sshgnsksncvvdvrvnvnkntdygnasnrxstannddnnnnyakkntnannnnnnnt	699		
Cp	597	TGTGGGTGCAGCAGCGCTGCCCTCTCCACAGAGTCTGTGCTGCACATACAGGCCAA	538		
Db	700	tgmnaadvysgnnnnnnnaarsngnyngndnsknvkvrgnrvnrvnsndrtnnn	759		
Cp	537	GAACACCATTCAGTACATCCCAAGAGAGAGTGTCCCGAGTAGAAGACAACCTGGCTC	478		
Db	760	nnvnnmrnwandnrdngnknrnnknngttsnndnnnmnyannnnknvnr	819		
Cp	477	GGTGACATCAGCAGGGTTTCCCTCTGCACCTTC-TCAGCTGCACATAGACCGAGCAGT	419		
Db	820	tnaynnnk'xanannynnnhsvannnkrgntvnandsvtnvnsdntansn	879		
Cp	418	TGACCCCGGCTCCAGTCGGGGTCGCAGACGGCTAGGAAGTTGGCT-TCAGACGCCCAATC	360		
Db	880	mnvvttnndnntcandndvkvntngdayvmyvsgngnrgnrgnrmarmnaand	939		
Cp	359	ATGTACTTGGCCAGCTGTGTCAAGACT-GGCTCAGCGCACGCCCAACAGGAGGTCCC	301		
Db	940	avsnrnrhrnhdnarnvgh	961		
Cp	300	CAGCACCTTGTATACAGCAGCT	279		

RESULTS 3

RESULTS

ID T23564 standard; cDNA to mRNA; 74 BP.

AC T23564;

DT 02-SEP-1996 (first entry)

DE Human gene signature HUMGS05416.

**KW** Gene signature; messenger RNA; mRNA; relative abundance; frequency; frequency;

KW	human; cloning; mapping; non-biased library; diagnosis; detection;
KW	cell typing; abnormal cell function; ss.
OS	Homo sapiens.
PN	WO9514772-A1.
PD	01-JUN-1995
PP	11-NOV-1994; J01916.
PR	12-NOV-1993; JP-353504.
PA	(MATS/) MATSUBARA K.
PA	(OKUB/) OKUBO K.
PI	Matsubara K, Okubo K;
DR	WPI: 95-206931/27.
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g
PT	for diagnosis of abnormal cell function, by preparing cDNA that
PT	reflects relative abundance of corresp. mRNA in specific human
PT	tissues

PS Claim 1; Page 1399; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

Query Match 5.7%; Score 70; DB 20; Length 74;

Best Local Similarity 95.9%;  
 Pred. No. 1.32e-22;  
 Matches 71; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 gatcanatagtgctgttttggtaaagtgaatgtatatntggttttaagtaaaataaggc 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 1113 GATCAGATAGTGCCTTTTGTAATAATGTAATGTATATGTGTTTTAGTAAATAGGC 1202

Db 61 acctqttcacaaa 74

50	01	acctgctccacaa	74
QY	1203	acctgtttcacaaa	1216

## 4 RESULT

ID Q51746 standard; cDNA; 91 BP.

AC Q51746;

DT 31-MAY-1994 (first entry)

DE Oligonucleotide probe MK14-A

KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;

OS Synthetic.

PN EP-571911-

PD 01-DEC-1993.

PF 24-MAY-1993;

PR 26-MAY-1992; US-8896.

PA (BECT ) BECTON DICKINSON

PI Shank DD, Spears PA;

DR WPI; 93-378844/48.

**New oligo:nucleotide probes specific for Mycobacteria - used for**

PT detection and amplification of *Mycobacteria* nucleic acid in samples

pt samples  
PS Claim 3: Page 14. 2300. English

PS Claim 3; Page 14; 23pp; English.

CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14

CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
CC (051735). It hybridized to all spp. of mycobacteria tested, but

CC (Q31/33): It hybridized to all spp. of mycobacteria tested. The probe may cross reacted to a few non-mycobacterial spp. CC

be useful as an initial screen for

CC see also Q51735-45 and Q51747-59.

See also Q51733, 43 and Q51747, 55.  
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.48; Score 42; DB 9; Length 91;



```

ID Q70467 standard; DNA; 114 BP.
AC Q70467;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PR 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65153.
DR Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)14(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 4.5%; Pred. No. 2.71e-05;
Matches 5; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
Qy 405 CTGAGCCGGGTCAACTGCTGCTATGTCAGCTGGAGAGGTGTGCAGGGGAACCC 464
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 114
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 465 TGTGTGATGTCACCGAGCCAGGTGTGCTTTCTACTCGGACACTCTTCCTTT 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.

```

```

OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PR 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65150 and R65151.
DR Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)6(TGC)(NNB)112(NNB)14(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-34. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.8%; Score 34; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 2.29e-04;
Matches 4; Conservative 33; Mismatches 75; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
Qy 405 CTGAGCCGGGTCAACTGCTGCTGCTATGTCAGCTGGAGAGGTGTGCAGGGGAACCC 464
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 114
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 465 TGTGTGATGTCACCGAGCCAGGTGTGCTTTCTACTCGGACACTCTTCCTTT 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
ID Q70466 standard; DNA; 114 BP.
AC Q70466;
DE 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR-9 peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"

```





CC sequence generates peptides that are cloverleaf in structure. Other  
CC generic sequences are shown in Q70455-68. Other specific peptides  
CC generated by these generic sequences are shown in R05150-34. TSARs are  
CC concatenated metofunctional proteins or peptides, comprising at least  
CC two functional regions - a binding domain with affinity for a ligand and  
CC a second effector peptide portion that is chemically or biologically  
CC active. They may further comprise a linker peptide between the 2 domains.  
CC The oligonucleotides are also designed so that the expressed peptide  
CC contains 2 or 4 cysteine residues positioned in, or flanking, the  
CC unpredicted or variant residues. These residues confer some degree of  
CC conformational rigidity to the peptides. The TSARs or compens. comprising



\*\*\*\*\*  
Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

MPsrch\_nnn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Thu Nov 6 12:46:31 1997; Maspar time 638.42 Seconds  
Tabular output not generated.  
Title: >US-08-842-827-7  
Description: (1-1232) from US08842827.seq  
Perfect score: 1232  
N.A. Sequence: 1 ACCATGCGAGCGAGCTGGGT.....CAAAAAAAAAAAAAAAAA 1232  
Comp: TGGTACGTGCGCTCCACCCA.....GTITTTTTTTTTTTTTTTTT  
Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0  
Searched: 707517 seqs, 256659390 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS  
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179  
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184  
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189  
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194  
195:EST195 196:EST196 197:EST197 198:EST198  
199:EST199

Statistics: Mean 11.465; Variance 3.184; scale 3.601

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	520	42.2	610	189	AA151344	z125a07.r1 Soares pre	0.00e+00
2	396	32.1	495	189	AA149244	z125a07.s1 Soares pre	0.00e+00
3	375	30.4	404	160	AA042850	zk52d08.r1 Soares pre	0.00e+00
4	339	27.5	349	105	HSC1GE101	H. sapiens partial cd	0.00e+00
5	297	24.1	413	21	R34980	yh86h03.r1 Homo sapie	0.00e+00
6	277	22.5	367	178	AA074521	zm17h03.s1 Stratagene	0.00e+00
7	255	20.7	264	105	HSC1GE102	H. sapiens partial cd	0.00e+00
8	195	15.8	237	21	R34878	yh86h03.s1 Homo sapie	3.44e-233
9	161	13.1	167	137	AA044400	zk52d08.s1 Soares pre	2.28e-183
10	162	13.1	177	162	AA102178	zm17h03.r1 Stratagene	7.99e-185
11	105	8.5	511	154	AA008988	mg99e04.r1 Soares mou	2.94e-103
12	97	7.9	423	78	R97295	yg74h06.r1 Homo sapie	3.95e-92
13	95	7.7	490	92	N31047	yx51h07.r1 Homo sapie	2.29e-89
14	75	6.1	446	29	R63796	y115h04.r1 Homo sapie	3.28e-62
15	72	5.8	472	114	W04958	zk43q09.r1 Soares fet	3.14e-58
16	65	5.3	357	159	AA037575	zk43f08.r1 Soares pre	4.59e-49
17	61	5.0	346	122	W61641	MD035all.r1 Life tech	4.48e-44
18	60	4.9	263	145	R75377	MD03538R Mouse brain,	1.22e-42
19	60	4.9	480	169	W45113	zc21c09.r1 Soares sen	1.22e-42
20	59	4.8	170	189	AA152122	z141g01.r1 Soares pre	2.29e-41
21	59	4.8	331	158	AA033777	zk19g11.r1 Soares pre	2.29e-41
22	59	4.8	398	138	AA048371	mk28d11.r1 Soares mou	2.29e-41
23	59	4.8	446	171	W67666	zk38a11.r1 Soares fet	2.29e-41
24	59	4.8	456	169	W45106	zc21a10.r1 Soares sen	2.29e-41
25	59	4.8	456	67	H68363	yr82f11.r1 Homo sapie	2.29e-41
26	59	4.8	471	189	AA152123	z141g02.r1 Soares pre	2.29e-41
27	59	4.8	474	113	W01275	za40f11.r1 Soares fet	2.29e-41
28	58	4.7	356	165	C17623	Human placenta CDNA 5	4.22e-40
29	57	4.6	348	54	H04659	yj49a08.r1 Homo sapie	7.71e-39
30	57	4.6	351	85	H57213	yr08d07.r1 Homo sapie	7.71e-39
31	57	4.6	385	88	H65929	yr86a11.r1 Homo sapie	7.71e-39
32	55	4.5	208	165	C16881	Human placenta CDNA 5	2.47e-36
33	55	4.5	449	173	W70040	zk49b06.s1 Soares fet	2.47e-36
34	55	4.5	451	83	H50533	yp08a11.s1 Homo sapie	2.47e-36
35	55	4.5	452	85	H54373	yq39g09.r1 Homo sapie	2.47e-36
36	55	4.5	555	161	AA055693	zk81b05.s1 Soares pre	2.47e-36
37	56	4.5	572	172	AA058383	zk81b05.r1 Soares pre	1.39e-37
38	54	4.4	545	137	AA043085	zk48b02.r1 Soares pre	4.35e-35
39	51	4.1	370	30	R71019	y150a11.r1 Homo sapie	2.17e-31
40	50	4.1	479	74	H90961	yu86e01.r1 Homo sapie	3.60e-30
41	50	4.1	485	60	H12562	yj12d10.r1 Homo sapie	3.60e-30
42	49	4.0	292	20	R00690	ye74a10.r1 Homo sapie	5.87e-29
43	47	3.8	422	4	T69858	yc18c09.r1 Homo sapie	1.49e-26
44	45	3.7	287	142	N86994	L1773F Fetal heart, L	3.51e-24
45	45	3.7	546	171	W67667	zk38a11.s1 Soares fet	3.51e-24

ALIGNMENTS

RESULT 1 AA151344 610 bp mRNA EST 10-DEC-1996  
LOCUS z125a07.r1 Soares pregnant uterus NbHPu Homo sapiens cDNA clone  
DEFINITION 502932 5'  
ACCESSION AA151344  
NID 91719889  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 610)

AUTHORS Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 452.

FEATURES  
Location/Qualifiers  
1..610  
/organism="Homo sapiens"  
/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAGAAATTCGCGCGCCCTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."  
/clone="502932"  
/clone.lib="Soares pregnant uterus NbHPU"  
/sex="female"  
/lab\_stage="adult"  
/lab\_host="DH10B"  
<1..>610

BASE COUNT 132 a 172 c 171 g 132 t 3 others

ORIGIN

Query Match 42.2%; Score 520; DB 189; Length 610;  
Best Local Similarity 96.4%; Pred. No. 0.00e+00;  
Matches 590; Conservative 0; Mismatches 12; Indels 10; Gaps 10;

Db 2 ttgcctctactgtgg -tacaccgctgtctgtattacaacaccactggagcgatgtcc 60  
Qy 620 TTGCCCTCTACGTGGCGTACACCCGCGTCTGTGATTAAACACCACTGAGCGATGTCC 679

Db 61 ttgttgccctctgaggggncactgggtggctgcctcactgtctgtcactctcagact 120  
Qy 680 TTGTTGGCCTCTCGAGGGGCGACTGGTGGCTGCCCTCACTGCTGCTACATCTCAGACT 739

Db 121 tcttcaaacccgaccccccacagcaactgtctgaaggaggagagctggaaacggaagccca 180  
Qy 740 TCTTCAAGCCCGACCCCCACAGCACTCTCTGAAGGAGGAGAGCTGGAAACCGGAACGCCA 799

Db 181 gcctgtcactgacgttgaccctggg -cgaggctgaccacaaccact -atgg -ataccgcg 237  
Qy 800 GCCTGTCACTGACGTTGACCCCTGGGCGAGGCTGACCACAACCACTTATGGGATACCCGC 859

Db 238 actctctctctgagggcgagcccgccacaggaaggaagctactgtggtccagctoa 297  
Qy 860 ACTCTCTCTCTGAGGCGGACCCCGCCCA -GGCA -GGAGAGTGTGTGTGATGCTCAGCTGA 917

Db 298 gggccaccacaggttggtccctccacagccctggttaggcactgagggctctggacgggtc 357  
Qy 918 TGCCCAACCA -GGTGGTCCCTCCA -GCC -TGGTATGACACTGAGGCTTCTGACGGGCTC 974

[illegible]

```

Query Match      32.1%; Score 396; DB 189; Length 495;
Best Local Similarity 93.8%; Pred. No. 0.00e+00;
Matches 467; Conservative 0; Mismatches 22; Indels 9; Gaps 9;

Db 1 gtgaacaggtgocctattttactaaaacacacatacatatttttcaaacagca 60
Cp 1213 GTGAACAGGTGCGCTATTTTACTAAAAACACATATACATTATTTACAAACAGCA 1154
Db 61 actatctgctctggtccctcttaacccataaaagaaggtatttgggacc 120
Cp 1153 ACTATCTGTCCTCGCTCCCTTCTTACCCCATAAAGAAGGGATATTTGGGACC 1094
Db 121 gacgggaacaggttccctcccaaatgctgaggggtaccagggtatctccagactcctggt 180
Cp 1093 GACGGGAACAGGTTCCCTCCCAATGCTGAGGGGTACCCAGGCATCTCCAGACTCTCTGTT 1034
Db 181 ccagtgcaggcaggcaggcagccgcctcactctccatcagccaggttccctgg 240
Cp 1033 CCAGTCAGGCGCAGGGGAGCGGAGACCGCTCACTGCTCCCATCAGCCAGGGTTCCTGG 974
Db 241 agccgtccagagccctcagtcgcttaaacacagggtgagggaccacacatgggtggccttc 300
Cp 973 AGCCGCTCCAGAACCTTCAGTGCCTAAACCAGG-CTGGAGGGACCACTGGGTGGCAT-C 916
Db 301 agctgactcacagtagctccctgctgctggggcggggtccgctcagagagagagtcgg 360
Cp 915 AGCTGACTCACAGCAGCTCCCTGCTGCGGGGGGGTCCGGCTCAGGAAGAAGAGTGGCG 856
Db 361 gttacc-ata-gtgggttggtgagctgcctcgcgc-agggtcaangtcangacagggctgg 417
Cp 855 GTATCCCATAAAGTGTGTGG-TCAGCCTCGCCCCAGGGTCAACGTCAAGTCAAGTCAAG 797
Db 418 gcttcggtttccagttcctctctttaagacagtgctgttgggttgggctttgaaga 477
Cp 796 GCTTCGGTT-CCAGCTCCCTCCCTCTTCAGACAGTCTGTGGGGT-CGGGCTTTGAAGAA 739
Db 478 ntttgagattntagcaga 495
Cp 738 GTCTGAGAT-GTAGCAGA 722

RESULT 3
LOCUS AA042850 404 bp mRNA EST 04-SEP-1996
DEFINITION zk53d08.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
ACCESSION AA042850
NID g1522366
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE The WashU-Merck EST Project
JOURNAL
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 237.
Location/Qualifiers

source 1..404
/organism="Homo sapiens"
/Note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACCTGGAGAATTCGCGCGCCCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/clone="486447"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
<1..>404 69 a 115 c 113 g 100 t 7 others

BASE COUNT
ORIGIN
Query Match 30.4%; Score 375; DB 160; Length 404;
Best Local Similarity 97.3%; Pred. No. 0.00e+00;
Matches 394; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

Db 1 acttctatagcgtctgcgaccccgactggagccgggtcaactgctcggtctatgtcagc 60
Cp 380 ACTTCTATAGCCTCTGCCACCCCGACTGGAGCGGGTCAACTGCTCGGTCTATGTGCAGC 439
Db 61 tggagaaggtgtgcaggggaaacccctnttgatgcacggagggccaggtttcttttact 120
Cp 440 TGGAGAAGGTGTGCAGGGGAAACCCCTGCTGATGTCACCGAGGCCAGGTTCTTTTACT 499
Db 121 cgggacactcttcccttgggagtactgcagtgtgtctcttggcgctgtatgtcagcagc 180
Cp 500 CGGGACACTCTTCCCTTGGGATGTACTGCATGGTGTCTTTGGCGCTGTATGTGCAGGCAC 559
Db 181 gactctgttggaaagtgggcagcggtgctgcgacccacagtcacantttctctggtggcct 240
Cp 560 GACTCTGTTGGAAGTGGGCACGCGTGTCTGCACCCACAGTCCAGTTCCTCTGGTGGCCT 619
Db 241 ttgccc-tctangtgggtacacccgcgtgtgtgtattacaacaccactggagagtgctc 299
Cp 620 TTGCCCTCTACCTGGGCTACACCCGCGTGTCTGATTACAAACACCACTGGAGCGGATGCC 679
Db 300 ttgttgctctctgcanggggcactgggtggctgcccctcaactgtctgtctacatctcagant 359
Cp 680 TTGTTGGCTCTCTGCAGGGGGCACGTGGTGGCTGCCCTCACTGTCTGCTACATCTCAGACT 739
Db 360 tcttnaaagcccccgcaccccccacaaagcactgtctnaagaggaggaga 404
Cp 740 TCTTCAAAGCCC-GACCCCCACA-GCACTGTCTGAAGGAGGAGGA 782

RESULT 4
LOCUS HSC1GE101 349 bp RNA EST 21-SEP-1995
DEFINITION H. sapiens partial cDNA sequence; clone c-1ge10.
ACCESSION Z43618
NID 9572809
KEYWORDS partial cDNA sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 349)
AUTHORS Genexpress.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE 2 (bases 1 to 349)
AUTHORS Genexpress.
TITLE The Genexpress cDNA program
JOURNAL Unpublished

```

3 (bases 1 to 349)  
 Aufrey, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,  
 Desnues, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B.,  
 Lorenzo, F., Mitchell, H., Marriage-Semson, R., Pietu, G., Pouillot, Y.,  
 Sebastiani-Kabakchis, C. and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534  
 Clone library from B. Soares, Psychiatry Dept. Columbia University  
 USA;

Cloning\_method: total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the lsmid BA  
 vector;  
 Sequencing\_method: single read, full automatic;  
 primer: M13-reverse  
 cDNA sequence colinear to mRNA  
 Stretch\_removed: nothing  
 Normalization\_method: Bento Soares, P.N.A.S in press;  
 Genexpress\_library\_id: C;  
 Genexpress\_sequence\_id: ylc-lgel0;

No significant homology found with :  
 genbank release 81 swissprot release 28.  
 Location/Qualifiers  
 1..349  
 /organism="Homo sapiens"  
 /isolate="muscular atrophy patient"  
 /dev\_stage="3 months old"  
 /tissue\_type="total brain"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"

BASE COUNT 56 a 108 c 97 g 82 t 6 others  
 ORIGIN

Query Match 27.5%; Score 339; DB 105; Length 349;  
 Best Local Similarity 98.0%; Pred. No. 0.00e+00;  
 Matches 340; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 3 accatgacggaggtgggtctctgctactgacgtgctggtgctgctgctgctgctcc 62  
 QY 1 ACATGACGGAGGTGGTCTTGGTCTGCTGACGTGCTGCTGCTGCTGCTGCTGCT 60  
 Db 63 ctgccccttgcctgctgacgtggtgaaagcccccgtacaaagcgaggtttactgcggg 122  
 QY 61 CTGCCCTTCGCTATCTGACGCTGGTGAAGCCCGGTACAAAGGAGGATTTACTGCGGG 120  
 Db 123 gatgactncaatcggtaccctaccgtccagttaccatcacccagggctgctgctggg 182  
 QY 121 GATGACTCCATCCGGTACCCCTACCGTCCAGATACCATACCCACGGGCTCATGCTGG 180  
 Db 183 gtcaccatcagggccacgctcatcttgtntcggcgggggaagcctaccgtgtgtacaca 242  
 QY 181 GTCAACCATCAGGGCCACCGTATCTTGTCTGCTCGCGGGGAAGCCCTACCTGGTGTACACA 240  
 Db 243 gacggctctattctgctggaactcaacaactagtgctgctgtatataaagtgctg 302  
 QY 241 GACCGGCTCTATCTGCTCGGACTTCAACACTAGTGGCTGCTGTATACAAAGTGCTG 300  
 Db 303 gggaccttctctgttggggctgcggtgagccagctctctgtgacagacct 349  
 QY 301 GGGACCTCTCTGTTGGGGCTGCGGTGAGCCAGTCTCTGACAGACCT 347

RESULT 5 R34980 413 bp mRNA EST 02-MAY-1995  
 LOCUS YN86H03.r1 Homo sapiens cDNA clone 136661 5'.  
 DEFINITION R34980  
 ACCESSION 9791881  
 NID EST.  
 KEYWORDS human clone-136661 library-Soares placenta Nb2HP vector-ptT7D3  
 SOURCE (Pharmacia) with a modified polylinker host=DH10B (ampicillin

resistant) primer-M13RP1 Raitel-Not I Reite2-Eco RI Female placenta  
 obtained at birth (full term). 1st strand cDNA was primed with a  
 Not I - oligo(dT) primer [5'  
 AATCGAAGAAATTCGGCGCGCAGGAATTTTTTTTTTTTTTTT 3'], double-stranded  
 cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the modified pT7n3  
 vector. Library went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens  
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 413)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, N., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

ORGANISM  
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 413)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, N., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

REFERENCE  
 AUTHORS  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevisan, E., Waterston, N., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

TITLE  
 JOURNAL  
 COMMENT

FEATURES  
 source  
 1..413  
 /organism="Homo sapiens"  
 /clone="136661"

BASE COUNT 74 a 119 c 120 g 93 t 7 others  
 ORIGIN

Query Match 24.1%; Score 297; DB 21; Length 413;  
 Best Local Similarity 95.5%; Pred. No. 0.00e+00;  
 Matches 338; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

Db 1 tccagttcttctggtggttgcctctacgtgggtacacccgctgtctgtattacca 60  
 QY 599 TCAGTTCTTCTGCTGGCTTTGCCCTCTAGTGGCTACACCCGCTGTCTGATTACA 658  
 Db 61 aacacacactgagcgatgctctgttggctctacaggggacactggtgctgctccca 120  
 QY 659 AACACACTGAGGAGATGCTCTGTGTGGCTCTGCGGGGACACTGTGTGCTGCCCTCA 718  
 Db 121 ctgtgtgtacatcctcagacttcttcaagcccgaccccccacacagcactgtgtgaaggag 180  
 QY 719 CTGTGTGTACATCTCAGACTTCTTCAAGCCCGACCCACACGACTGTCTGAAGGAGG 778  
 Db 181 aggagctggaagcgaagcccgacgtctacactgacgtgacccctggcgagggcgacac 240  
 QY 779 AGGAGCTGGAAGCAAGCCAGCTGTCTGACGTGTGACCTGTGGGGGAGCGTGACCA 838  
 Db 241 aaccactatgggtacaccccgacactctctctctgagggccggagcccccagagggg 300  
 QY 839 AACCACTTATGGGATACCGCACTTCTTCTCTGAGG-CCGGACCCCGCCGAGGAGG- 896  
 Db 301 agctgctgttgaattcagttttagggccacccagctgttcttcttctcagctng 354  
 QY 897 AGCTGCTCT-GAGT-CCAGCTGATG-CCACCCAGGTGTCTCCT-CCAGCCTGG 946

RESULT 6 AA074521 367 bp mRNA EST 07-OCT-1996  
 LOCUS zmi7h03.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
 DEFINITION 525941 3'.  
 ACCESSION AA074521  
 NID g1614408

```

KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
REFERENCE Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 367)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maria,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 194.

FEATURES
source
1..367
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'
CTCGAGTCTTTTATTTTATTTTATTTT 3'"
/clone="525941"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
complement(1..>367)
89 a 107 c 97 g 70 t 4 others

BASE COUNT
ORIGIN
Query Match 22.5%; Score 277; DB 178; Length 367;
Best Local Similarity 96.2%; Pred. No. 0.00e+00;
Matches 328; Conservative 0; Mismatches 5; Indels 8; Gaps 7;

Db 1 tgtgaacagggtgcctattttactaaaccacacatacatattttacaaacagc 60
|||||
Cp 1214 TGTGAACAGGTGCGCCTATTTTACTAAACCCACATATACATTATTTACAAACAGC 1155
|||||
Db 61 aactatctgctctcggtccctcttaccctcaaaagaggaggtatttgggac 120
|||||
Cp 1154 AACTATCTGATCTCTGCTCCCTCTTAACCCCATATAAAGAGGGGATATTGGGGAC 1095
|||||
Db 121 cgacgggaacaggttccctcccaaatctgagggtaccagggaatctccagactctg 180
|||||
Cp 1094 CGACGGGAACAGGTTCGCCCTCCAAATGCTGAGGGGTACCGAGCATCTCCAGACTCTCG 1035
|||||
Db 181 tccagtcagggcagggcgagcgactctcactctccatccatccagcaggttcc 240
|||||
Cp 1034 TCCAGTCGAGGGCAGGGGGCAGCGGAGACC-GTCACTGCTCCATCAGCCCGGGTTCCT 976
|||||
Db 241 ggaagccgtccagaagccctcagtcgctaancaaggggtggaaggaccactngtgg 300
|||||
Cp 975 GGAGCCGCTCCAGAA-CCCTCAGTGGCTTAACCA--GGCTGGA-GGGACCACTGGGTGG 920
|||||
Db 301 cctaagctgggactcacagcagctcccttgcctgtagggcgg 341
|||||
Cp 919 CATCA-GCTGG-ACTCACACAGCTCCCT-GCCTGGGGGGG 882
|||||

RESULT 7 HSC1GE102 264 bp RNA EST
LOCUS H. sapiens partial cdna sequence; clone c-1ge10. 21-SEP-1995
DEFINITION
ACCESSION Z39680

```

---

```

NID 955475
KEYWORDS Partial cdna sequence; transcribed sequence fragment.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 264)
Direct Submission
Genexpress.
TITLE Submitted (24-OCT-1994) Genethon, B.P. 60, 91002 Evry Cedex France
JOURNAL and Genetique Moleculaire et Biologie du developpement, CNRS UPR420
B.P. 8, 94801 Villejuif Cedex France.E-mail: genexpress@genethon.fr
REFERENCE 2 (bases 1 to 264)
Genexpress.
AUTHORS The Genexpress cDNA program
TITLE Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 264)
Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Clone library from B.Soaress, Psychiatry Dept. Columbia University
USA;

Cloning_method: total mRNA was oligo-(dt) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA
vector;
Sequencing_method: single read, full automatic;
Primer: (-21)M13 universal;
cDNA_sequence_complementary to mRNA (3'end)
Stretch_removed: 18 T removed at sequence 5'end
Normalization_method: Bento Soares, P.N.A.S in press;
Genexpress_library_id: C;
Genexpress_sequence_id: alc-1ge10;

No significant homology found with :
genbank release 81 swissprot release 28.

FEATURES
source
1..264
/organism="Homo sapiens"
/isolate="muscular atrophy patient"
/dev_stage="3 months old"
/tissue_type="total brain"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
BASE COUNT 66 a 79 c 61 g 54 t 4 others
ORIGIN
Query Match 20.7%; Score 255; DB 105; Length 264;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 256; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 4 aaagggtgcctattttnaataaaacacatacatattttacaaacagcaacta 63
|||||
Cp 1209 AACAGGTGCGCCTATTTTACTAAACCCACATATACATTATTTTACAAACAGCAACTA 1150
|||||
Db 64 tctgactctcgtctcccttcttaccctcaaaagagggtatatttggggaccgac 123
|||||
Cp 1149 TCTGATCTCTCGGTCCTCTTCTTAACCCCATATAAAGAGGGATATTTTGGGACCAGC 1090
|||||
Db 124 ggaacaggttccctcccaaatctgagggtaccaggcatctccagactctggtccag 183
|||||
Cp 1089 GGAACAGGTTCCTCCCAATGCTGAGGGCTACCCAGGCATCTCCAGACTCCTGGTCCAG 1030
|||||
Db 184 tgcagggcagggggcagcganccgntcactgtctcccatcagccaggggttctctggagcc 243
|||||
Cp 1029 TGCAGGCAGGGGGCAGCGGACCGCTCACTGCTCCCATCAGCCAGCGGTTCCTGGAGCC 970
|||||
Db 244 cgtccagagccctcagtgacct 264

```

CP	969	CGTCCAGAACCCCTCAGTGCCT	949
RESULT	B		
LOCUS	R34878	237 bp	MRNA
DEFINITION	YH69h03.s1 Homo sapiens cDNA clone 136661 3'.	EST	02-MAY-1995
ACCESSION	R34878		
NID	9791779		
KEYWORDS	EST.		
SOURCE	human clone-136661 library=Soares placenta NB2HP vector=PT7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=2lm3 Rstai=Not I Rstai2=ECO RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAAATTCGGCGCGCAGGAATTTTTTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified PT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaudo.		
ORGANISM	Homo sapiens		
REFERENCE	Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 237) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 151 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.		
FEATURES	Location/Qualifiers 1..237 /organism="Homo sapiens" /clone="136661"		
BASE COUNT	62 a 65 c 53 g 50 t		7 others
ORIGIN			
Query Match	15.8%	Score 195;	DB 21;
Best Local Similarity	95.1%	Pred. No. 3,44e-233;	
Matches	215;	Conservative	0; Mismatches 8; Indels 3; Gaps 3;
Db	12	tntttactaaaaaacacatacatattacattttacaaaacagcaactatctgctctc	71
CP	1198	TATTTTACTTAAAAACCATATACATACATTTTACAAAACAGCAACTATCTGATCTCTC	1139
Db	72	ggncccttcttaacccataaaaaagaggggatatttgggacacgagggcaacaggttc	131
CP	1138	GGTCCCTTCTCTTAACCCCATAAAAGAGGGGATATTTGGGGACCGCGGGAACAGGTTC	1079
Db	132	cctctcaaatctgtagggctaccaggggnatctccagactctctgggnccagtcgagggca	191
CP	1078	CCTCTCAAAATGCTAGGGCTACCCAGG-CATCTCCAGACTCTCTGG-TCCAGTGCAGGGCA	1021
Db	192	gggggagcgagcgctccactgntcccatcagccagcgggntct	237
CP	1020	GGGGCAGCGGAACCGCTC-ACTGCTCCCATCAGCCAGGGTCTCT	976
RESULT	9		
LOCUS	AA044400	167 bp	MRNA
			EST
			04-SEP-1996

DEFINITION	2k5d08.sl Soares pregnant uterus NbHPU Homo sapiens cDNA clone 48647 3'.					
ACCESSION	AA044400					
NID	91522275					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrates; Eutheria; Primates; Catarhini; Hominiidae; Homo. 1 (bases 1 to 167)					
AUTHORS	Hillier,L., Clark,M., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares.M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson.R.					
TITLE	The WashU-Merck EST Project					
JOURNAL	Unpublished (1995)					
COMMENT	 Contact: Wilson RK WashU-Merck EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu  This clone is available royalty-free through LINL ; contact the IMAGE Consortium ( <a href="mailto:info@image.lnl.gov">info@image.lnl.gov</a> ) for further information. Seq primer: -40M13 fwd. from Amerham High quality sequence stop: 144. Location/Qualifiers 1..167 /organism="Homo sapiens" /note="Organ: uterus; Vector: pTn3-Pac; Site_1: Not I; Site_2: Eco RI; lsc strand cDNA was primed with a Not I - Oligo(dT) primer [5' AATCGGAAGAAATTCCGGGCGCCTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo." /clone="486447" /clone_lib="Soares pregnant uterus NbHPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" BASE COUNT     46 a     37 C     24 g     60 t ORIGIN					
	Query Match         13.1%; Score 161; DB 137; Length 167;					
	Best Local Similarity 100.0%; Pred. No. 2.28e-183;					
	Matches     161; Conservative     0; Mismatches     0; Indels     0; Gaps     0;					
Ddb	7   tttttttttttttttttggaaacaggcgccctattttactaaaaaccacatatcat 66 					
Cp	1232 TTTTFTTTTTTTTTTTTTTGTAACAAGTCGCCTATTTACTAAAACCACATATACAT 1173 					
Ddb	67 tacatttcacaagaagcaactatcgtatctcgcgtccctccaaccocataaaaag 126 					
Cp	1172 TACATTTTCAAAACAGCAACTATCTCATCTCTCGGTGCCCTTCCTTAACCCCATAAAAAAG 1113 					
Ddb	127 aaagggatatttgaggaccgacgggaacagggtcccctoca 167 					
Cp	1112 AAGGGGATATTGGGGACCGACGGGAACAGGTCCCGCTCCA 1072 					
RESULT	10					
LOCUS	AA102178     177 bp     mRNA     EST     28-OCT-1996					
DEFINITION	zml7n03.r1 Stratiogene pancreas (#377208) Homo sapiens cDNA clone 525941 5'.					
ACCESSION	AA102178					
NID	91646351					









```
QY 445 AGGTGTCAGGGGAACCCGCTGATGTCACCGAGGCCAGGTGCTTTCTACTCGGGAC 505
Db 345 actotctgttttccatgtactgcatgctgtttgtggcacttttatcttcaagccaggatga 405
QY 505 ACTCTTCCTTTGGGATGTACTGTCATGGTGTCTTGGCGCTGTATGTGCAGGCACGACTCT 565
Db 405 agggagactggggaagactcttacgccccacactgcaatt 445
QY 565 GTTGGAACTGGGCACGGCTGCTGGGACCCACAGTCCAGTT 605
```

Search completed: Thu Nov 6 12:58:09 1997  
Job time : 698 secs.

\*\*\*\*\*

MAPSRLH

(TM)

\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

MPSrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Nov 6 12:58:35 1997; MasPar time 336.04 Seconds  
Tabular output not generated. 1007.529 Million cell updates/sec

Title: >US-08-842-827-7

Description: (1-1232) from US08842827.seq

Perfect Score: 1232

N.A. Sequence: 1 ACCATGCCACGGAGTGCGGT.....CAAAAAAAAAAAAAAAAAA 1232

Comp: TGGTAGTGCCTCCACCCA.....GTTTTTTTTTTTTTTTTT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-STS-THREE

1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204  
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210  
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215  
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220  
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225  
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230  
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3  
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10  
46:STS11 47:STS12 48:STS13

Database: EST-STS-FOUR

49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5  
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10  
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15  
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20  
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25  
74:gnSTS 75:enEST1 76:enEST2 77:enEST3 78:enEST4  
79:enEST5 80:enEST6 81:enEST7 82:enEST8 83:enEST9  
84:enEST10 85:enEST11 86:enEST12 87:enEST13 88:enEST14  
89:enEST15 90:enEST16 91:enEST17 92:enEST18 93:enEST19  
94:enEST20 95:enEST21 96:enEST22 97:enEST23 98:enEST24  
99:enEST25 100:enEST26 101:enEST27 102:enEST28  
103:enEST29 104:enEST30 105:enEST31 106:enSTS 107:ueEST1  
108:ueEST2

Statistics: Mean 11.492; Variance 3.475; scale 3.307

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
C 1	458	37.2	491	25	N54789	YV31B01.s1 Soares fet	0.00e+00
2	395	32.1	478	26	N75714	YV31B01.r1 Soares fet	0.00e+00
3	66	5.4	385	33	AA107326	ML95B10.r1 Stratagene	3.53e-46
4	61	5.0	337	27	W39816	309 Mouse VM cDNA lib	2.96e-40
5	60	4.9	402	27	W39811	307 Mouse VM cDNA lib	4.40e-39
6	59	4.8	564	88	W30942	ZC64F09.r1 Soares fet	6.47e-38
7	59	4.8	564	88	HS942332	ZC64F09.r1 Soares fet	6.47e-38
8	56	4.5	247	33	W39815	308 Mouse VM cDNA lib	1.94e-34
9	47	3.8	377	33	AA106892	ML85A07.r1 Stratagene	2.62e-24
C 10	42	3.4	313	51	MM1155742	MX78G01.r1 Soares mou	6.44e-19
C 11	42	3.4	313	66	AA237956	MX78G01.r1 Soares mou	6.44e-19
C 12	41	3.3	415	87	HS605336	ZB63F08.r1 Soares fet	7.27e-18
C 13	34	2.8	914	34	AA141702	CK02248.contig Drosop	8.73e-11
C 14	32	2.6	422	34	AA142082	CK02216.3prime Drosop	7.12e-09
C 15	30	2.4	446	12	AA181403	ZP95F10.s1 Stratagene	5.05e-07
C 16	28	2.3	228	87	HS50335	ZB91E04.s1 Soares par	3.05e-05
C 17	28	2.3	236	81	HS1146241	ZP52F10.r1 Stratagene	3.05e-05
C 18	28	2.3	263	12	AA182434	ZP52F10.r1 Stratagene	3.05e-05
C 19	28	2.3	263	77	AT6391	2647 Arabidopsis thal	3.05e-05
C 20	28	2.3	343	63	AA225572	NC08C05.s1 NCI CGAP P	3.05e-05
C 21	28	2.3	343	81	HS1144996	NC08C05.s1 NCI CGAP P	3.05e-05
C 22	28	2.3	435	14	AA187745	ZP72G10.s1 Stratagene	3.05e-05
C 23	28	2.3	495	11	AA176159	ZP23H04.s1 Stratagene	3.05e-05
C 24	27	2.2	100	6	AA120007	MM15H06.r1 Beddington	2.22e-04
C 25	27	2.2	252	99	MAA31554	MX31F08.r1 Soares mou	2.22e-04
C 26	27	2.2	252	64	AA231554	MX31F08.r1 Soares mou	2.22e-04
C 27	27	2.2	260	101	MAA60919	MX77H08.r1 Soares mou	2.22e-04
C 28	27	2.2	336	91	MM1156654	MX92D02.r1 Soares mou	2.22e-04
C 29	27	2.2	336	67	AA238805	MX92D02.r1 Soares mou	2.22e-04
C 30	27	2.2	347	39	G09147	human STS CHLC.GATA2B	2.22e-04
C 31	27	2.2	351	101	MAA60235	VA55D02.r1 Soares mou	2.22e-04
C 32	27	2.2	385	78	AT9619	10739 Arabidopsis tha	2.22e-04
C 33	27	2.2	395	3	AA112233	ZM64B11.s1 Stratagene	2.22e-04
C 34	27	2.2	403	87	HS733324	ZC64F09.s1 Soares fet	2.22e-04
C 35	27	2.2	403	1	W02733	ZC64F09.s1 Soares fet	2.22e-04
C 36	27	2.2	405	6	AA121078	ZM22B08.r1 Stratagene	2.22e-04
C 37	27	2.2	432	99	MAA37207	MX18E02.r1 Soares mou	2.22e-04
C 38	27	2.2	432	66	AA237207	MX18E02.r1 Soares mou	2.22e-04
C 39	27	2.2	435	33	AA105273	ML82H08.r1 Stratagene	2.22e-04
C 40	26	2.1	187	102	MAA44695	MT13B09.r1 Soares mou	1.54e-03
C 41	26	2.1	210	84	HS1154328	ZS09E01.s1 Soares NDH	1.54e-03
C 42	26	2.1	335	67	AA238825	MX92G02.r1 Soares mou	1.54e-03
C 43	26	2.1	335	91	MM1156666	MX92G02.r1 Soares mou	1.54e-03
C 44	26	2.1	343	5	AA117600	MP69A04.r1 Soares 2Nb	1.54e-03
C 45	26	2.1	461	30	AA063613	ZE87F05.s1 Soares fet	1.54e-03

ALIGNMENTS

RESULT 1  
LOCUS YV31B01.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone  
DEFINITION 491 bp mRNA EST 28-JAN-1997  
244297 3'  
ACCESSION N54789  
NID G1196109  
KEYWORDS EST.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,  
Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine





```

Db 121 tggccttcacacgggattgtcacgggtatctgaactacaagcatcctcctagcagtcc 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 620 TTGCCCTCTAGTGGGTACACCGCGGTGCTGATTACAAACACCACTGGAGCGATGCC 679
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 181 tggcaggtatggccaaaggagctgtgtgacctgctgcatagtggttcttcgtgcgacc 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 680 TTGTTGGCTCTCTGACGGGGACATGTGGTGCCTCCTCACTGCTGCTACATCTCAGACT 739
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 241 tcttcaa 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 740 TCTTCAA 745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LOCUS W39811 402 bp mRNA EST 05-FEB-1997
DEFINITION 307 Mouse VM cDNA library Mus musculus cDNA clone smp22 1.22.
ACCESSION W39811
NID g1816977
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 402)
AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.
TITLE Sequence analysis of 497 mouse brain ESTs expressed in the
JOURNAL Genomics 39, 147-153 (1997)
COMMENT
Contact: Davies,R.W.
Robertson Laboratory of Biotechnology
Institute of Biomedical and Life Sciences, Division of Molecular
Genetics
Robertson Building, 54 Dumbarton Road, Glasgow G11 6NU
Tel: 44 141 330 5102
Fax: 44 141 330 5102/4878
Email: gbgazl@udcf.gla.ac.uk
Seq primer: T7/T3alpha
High quality sequence stop: 422.
Location/Qualifiers
1..402
/organism="Mus musculus"
/clone="smp22 1.22"
/clone_lib="Mouse VM cDNA library"
/tissue_type="ventral midbrain"
<1..>402

BASE COUNT 69 a 131 c 101 g 101 t
ORIGIN

Query Match 4.9%; Score 60; DB 27; Length 402;
Best Local Similarity 62.2%; Pred. No. 4.40e-39;
Matches 153; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Db 1 gggcacgcctctctccatgttcacatgctgctgtatctgtggtcgtcacttcaggccgg 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 501 GGGACACTCTCTCTTGGGATGTACTGCATGGTGTCTTGGGCGCTGTATGTGCAGGCACG 560
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 61 ctccacctggcgggggccgactgctcccccctcccgacgttcaacttttgcctcatgat 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 ACTCTGTTGGAAGTGGGCACGGCTGCTGCACCCACAGTCCAGTCTTCTCTGTTGGCCTT 620
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 121 ggccttcacacgggattgtcacgggtatctgactacacgcatcatcctcagcagtctct 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 621 TGCCCTCTACGTGGGTACACCCCGTGTCTGTGATTACAAACCACTGGAGCGATGCTCT 680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

; Db 181 ggcaggatttggccaaaggagctgtgtggcctgctgcatagtggttcttcgtgtccgacct 240

```

```

QY 681 TGTGGCCTCTCTGAGGGGCACTGGTGGCTGCCCTCTACTGTCTGCTACATCTCAGACTT 740
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ctccaa 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 741 CTTCAA 746
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
LOCUS W30942 564 bp mRNA EST 25-NOV-1996
DEFINITION z64f09.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
327113 5' similar to WP:T28D9.3 CE02068 ;.
ACCESSION W30942
NID g1311934
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 564)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 774 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 336.
Location/Qualifiers
1..564
/organism="Homo sapiens"
/clone="327113"
/clone_lib="Soares fetal heart NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
<1..>564

BASE COUNT 128 a 155 c 139 g 134 t 8 others
ORIGIN

Query Match 4.8%; Score 59; DB 1; Length 564;
Best Local Similarity 60.9%; Pred. No. 6.47e-38;
Matches 207; Conservative 0; Mismatches 129; Indels 4; Gaps 4;

Db 192 aaccctcactggtggcagcactcctataagcagtggtgctcttctctcttgggtgcga-tc 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 AACACTACGTGGTGTGTATACAGGTCGTGGGACCTTCTCTGTTGGGCTGCCGTG 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

; Db 251 agccagtctttcacagacattgccaaagtccatagggcgctgcctccctactcttg 310

```



QY 328 AGCCAGTCTCTGACAGACCTGCGCAAGTACATGATGGGGCTCTGAAGCCCAACTTCTCT - 386  
 Db 311 aagtgtctcaacctgtatttcagcagatcaactgtctgaaggctacattcagaacta 370  
 QY 387 AGCCGTCCTGGACCCCGACTGGAGCGGGTCACTGCTCGGTCTATGTGCAGCTGGAGAA 446  
 Db 371 cagatcagaggtgatgacagcaaaagtcagcagcagcaggaagtcccttctcttgcca 430  
 QY 447 GGTGTGACAGGGGAAACCTGCTGATGTACCGAGGCCAGGTTGTCTTCTACTCGGGACA 506  
 Db 431 tgcctctctccatgataactgtctgaattgggtgctatantcagagcccgcttca 490  
 QY 507 CTCTCTCTTGGAGTACTGATGCTGTTCTT-GGCGCTGTATGTGCAGGCAGACTCT 565  
 Db 491 cttggcagagc-cgncgtgctcggccctctcgcagtt 529  
 QY 566 GTTGAAGTGGGACGCTGCTGCGACCCACAGTCCAGTT 605  
 RESULT 7  
 ID HS942332 standard; RNA; EST; 564 BP.  
 AC W30942;  
 NI g1311934  
 DT 13-MAY-1996 (Rel. 47, Created)  
 DT 07-MAR-1997 (Rel. 51, Last updated, Version 2)  
 DE zc64f09.r1 Soares fetal heart NBHL19W homo sapiens cDNA clone  
 DE 327113 5' similar to WP-T28D9.3 CE02068 ;  
 KW EST.  
 OS Homo sapiens (human)  
 OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 OC Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP 1-564  
 RA Hüllner L., Clark N., Dubucque T., Elliston K., Hawkins M.,  
 RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
 RA Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F.,  
 RA Trevasakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;  
 RT "The WashU-Merck EST Project";  
 RL Unpublished.  
 CC Contact: Wilson RK WashU-Merck EST Project Washington University  
 CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,  
 CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:  
 CC est@wustl.edu This clone is available royalty-free through  
 CC LUNL; contact the IMAGE Consortium (info@image.lnl.gov) for  
 CC further information. Insert Length: 774 Std Error: 0.00 Seq primer:  
 CC mob.REGA+ET High quality sequence stop: 336.  
 FH Key Location/Qualifiers  
 FT source 1..564  
 FT /organism="Homo sapiens"  
 FT /note="Organ: heart; Vector: pT7n3D (Pharmacia) with a  
 FT modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 FT strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 FT TCTTACCACTGAGTGGAGCGCGCATCTTTTTTTTTTTT 3']  
 FT double-stranded cDNA was size selected, ligated to Eco RI  
 FT adapters (Pharmacia), digested with Not I and cloned into  
 FT the Not I and Eco RI sites of a modified pT7n3 vector  
 FT (Pharmacia). Library went through one round of  
 FT normalization to a Cot = 5. Library constructed by M.Fatim  
 FT a  
 FT Ronaldo. This library was constructed from the same fetus  
 FT as the fetal lung library, Soares fetal lung NBHL19W."  
 FT /clone="327113"  
 FT /clone\_lib="Soares fetal heart NBHL19W"  
 FT /sex="unknown"  
 FT /dev\_stage="19 weeks"  
 FT /lab\_host="DH10B (ampicillin resistant)"  
 FT <1..564  
 FT mRNA  
 SQ Sequence 564 BP; 128 A; 155 C; 139 G; 134 T; 8 other;  
 Query Match 4.8%; Score 59; DB 88; Length 564;  
 Best Local Similarity 60.9%; Pred. No. 6.47e-38;

Matches 207; Conservative 0; Mismatches 129; Indels 4; Gaps 4;  
 Db 192 aaccctcaagtggaagcactctataagaagtgagggtgtctctcttttggtgca-tc 250  
 QY 268 AACAACTAGTGGCTGCTGTATACAGGTGCTGGGGACCTCTCTGTTTGGGCTGCCGTG 327  
 Db 251 agcagctcttccagacattgccaaagtgtccatagggcgctgctcactctcttg 310  
 QY 328 AGCCAGTCTCTGACAGACTGCGCAAGTACATGATTTGGGGTCTGAAGCCCAACTTCTCT - 386  
 Db 311 aagtgtctcaacctgtatttcagcagatcaactgtctgaaggctacattcagaacta 370  
 QY 387 AGCCGTCCTGGACCCCGACTGGAGCGGGTCACTGCTCGGTCTATGTGCAGCTGGAGAA 446  
 Db 371 cagatcagaggtgatgacagcaaaagtcagcagcagcaggaagtcccttctcttgcca 430  
 QY 447 GGTGTGACAGGGGAAACCTGCTGATGTACCGAGGCCAGGTTGTCTTCTACTCGGGACA 506  
 Db 431 tgcctctctccatgataactgtctgaattgggtgctatantcagagcccgcttca 490  
 QY 507 CTCTCTCTTGGAGTACTGATGCTGTTCTT-GGCGCTGTATGTGCAGGCAGACTCT 565  
 Db 491 cttggcagagc-cgncgtgctcggccctctcgcagtt 529  
 QY 566 GTTGAAGTGGGACGCTGCTGCGACCCACAGTCCAGTT 605  
 RESULT 8  
 LOCUS W39815 242 bp mRNA EST 05-FEB-1997  
 DEFINITION 308 Mouse VM cDNA library Mus musculus cDNA clone spmp22 1.23.  
 ACCESSION W39815  
 NID g1816978  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Mus.  
 REFERENCE 1 (bases 1 to 242)  
 AUTHORS Stewart,G.J., Savioz,A. and Davies,R.W.  
 TITLE Sequence analysis of 497 mouse brain ESTs expressed in the  
 JOURNAL substantia nigra  
 COMMENT Genomics 39, 147-153 (1997)  
 CONTACT: Davies,R.W.  
 Roberton Laboratory of Biotechnology  
 Institute of Biomedical and Life Sciences, Division of Molecular  
 Genetics  
 Roberton Building, 54 Dumbarton Road, Glasgow G11 6NU  
 Tel: 44 141 330 5102  
 Fax: 44 141 330 5102/4878  
 Email: gbg21@udcf.gla.ac.uk  
 Seq primer: 17/T3alpha  
 High quality sequence stop: 2.  
 FEATURES  
 source 1..242  
 /organism="Mus musculus"  
 /note="Organ: brain; Vector: pSPORT1; Mouse Ventral  
 Midbrain directional cDNA library in pSPORT1. The library  
 was created by subtractive hybridisation using VM mRNA as  
 the target nucleic acid population and total cerebellar ss  
 cDNA as the driver nucleic acid population. The clones are  
 3' directed."  
 /clone="spmp22 1.23"  
 /clone\_lib="Mouse VM cDNA library"  
 /tissue\_type="ventral midbrain"  
 <1..242  
 mRNA 34 a 82 c 60 g 66 t  
 BASE COUNT  
 ORIGIN  
 Query Match 4.5%; Score 56; DB 27; Length 242;  
 Best Local Similarity 61.6%; Pred. No. 1.94e-34;  
 Matches 149; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```

Db 1 cggggcagcgcctcttccactgttctcactatgtctgtatgtgtgtctctacattcaggccc 60
QY 500 CCGGACACTCTCTTGGGATGACGATGTGTCTTGCGGTGTATGTGCGAGGCAC 559
Db 61 gcttcacctggcgggccgagctgtccgcccctctctcagttgcacttttgcattga 120
QY 560 GACTCTGTGGAAGTGGGACGGGTGTCGACCCACAGTCCAGTTCCTCGTGGGCGCT 619
Db 121 tggccttctacacggattgttcaagggtatctgaactacaagcatcctcagcgtgtcc 180
QY 620 TTGCCCTCTAGTGGGCTACACCGCGCTGTGATTACAAACACCCAGGAGCGATGCC 679
Db 181 tggcaggatttgcacaggagctgtgtgcccgtgctgcatagtgttcttctgtccgacc 240
QY 680 TTGTGGCTCTCTGACGGGCGCATGTGTGCTGCCCTCAGTCTGCTACATCTCAGACT 739
Db 241 tc 242
QY 740 TC 741

RESULT 9 AA106892 87 bp mRNA EST 04-FEB-1997
LOCUS ml85a07.r1 stratagene mouse kidney (#937315) Mus musculus cDNA
DEFINITION clone 518772 5' similar to TR:G1161100 G1161100 HYDROGEN
PEROXIDE-INDUCIBLE PROTEIN ;
ACCESSION AA106892
NID AA106892
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 87)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:312620
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 55.
Location/Qualifiers
1..87
/organism="Mus musculus"
/strain="C57/B16"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTGGAGTTTATTTTATTTT 3'"
/clone="518772"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/dev_host="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
<1..>87

mRNA
BASE COUNT 18 a 21 c 20 g 28 t

ORIGIN
Query Match 3.8%; Score 47; DB 33; Length 87;
Best Local Similarity 80.5%; Pred. No. 2.62e-24;
Matches 62; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 5 gtcaagagggcagggtgtcttcttactcggacactcttctctatgtactatgcag 64
QY 472 GTCACCGAGGCCAGGTGTCTTCTACTCGGACACTCTTCTTGGGATGTACTGCGAT 531
Db 65 ctgtgtgtcgcacttta 81
QY 532 GTGTCTTGTGGCGCTGTA 548

RESULT 10
ID M01155742 standard; RNA; EST; 313 BP.
AC AA237956;
NI G1862038
DT 06-MAR-1997 (Rel. 51, Last updated, Version 1)
DE mx78g01.r1 Soares mouse NML Mus musculus cDNA clone 692496 5'
DE similar to TR:G1161100 G1161100 HYDROGEN PEROXIDE-INDUCIBLE PROTEIN
DE :
DE EST.
KW Mus musculus (house mouse)
OS Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC Mus.
OC [1]
RN [1]
RP 1-313
RA Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N.,
RA Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J.,
RA Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K.,
RA Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R.,
RA Waterston,R.;
RA "The WashU-HMI Mouse EST Project";
RT Unpublished.
RL Unpublished.
CC Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project
CC Washington University School of Medicine 4444 Forest Park Parkway,
CC Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810
CC Email: mouseest@wustl.wustl.edu This clone is available
CC royalty-free through LNL ; contact the IMAGE Consortium
CC (info@image.llnl.gov) for further information. MGI:426036 Possible
CC reversed clone: similarity on wrong strand Seq primer: -28m13 rev2
CC ET from Amersham High quality sequence stop: 57.
FH Key Location/Qualifiers
FT source 1..313
FT /organism="Mus musculus"
FT /notes="vector: pT73D-Pac (Pharmacia) with a modified
FT polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
FT was primed with a Not I - oligo(dt) primer [5',
FT TGTACCAATCTGAAGTGGGCGCGCGAATCTTTTATTTT 3'] ;
FT double-stranded cDNA was ligated to Eco RI adaptors
FT (Pharmacia), digested with Not I and cloned into the Not I
FT and Eco RI sites of the modified pT73 vector. Library
FT constructed and normalized by Bento Soares and M.Fatima
FT Bonaldo."
FT /clone="692496"
FT /clone_lib="Soares mouse NML"
FT /tissue_type="Liver"
FT /lab_host="DH10B"
FT <1..>313
FT mRNA
SQ Sequence 313 BP; 83 A; 58 C; 85 G; 87 T; 0 other;

Query Match 3.4%; Score 42; DB 91; Length 313;
Best Local Similarity 69.4%; Pred. No. 6.44e-19;
Matches 84; Conservative 0; Mismatches 36; Indels 1; Gaps 1;

Db 2 ggagtcacactgtgacgctcactcagtggtgtttagtcagacactcagaa-aggcca 60
QY 691 GGAGGCCAACAGGACATCGCTCCAGTGGTGTGTATACACACGCGGTGTAGCCCA 632

```



LOCUS AA141702 914 bp mRNA EST 25-JAN-1997  
 DEFINITION CK02248.contig Drosophila Embryonic Polysomal in pBS skt Drosophila melanogaster cDNA clone CK02248 complete.  
 ACCESSION AA141702  
 NID 91705147  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 914)  
 AUTHORS Kopczyński, C., Serrano, T., Rubin, G. and Goodman, C.  
 TITLE mRNAs enriched for membrane and secreted proteins versus cytosolic proteins  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Lewis, S.  
 G. M. Rubin-Molecular and Cell Biology  
 University of Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: ES@fruitfly.berkeley.edu, http://fruitfly.berkeley.edu/  
 Insert Length: 914 Std Error: 0.00  
 Plate: 8 row: 3 column: 1  
 FEATURES  
 Location/Qualifiers  
 1..914  
 /organism="Drosophila melanogaster"  
 /note="Organ: embryo; Vector: plasmid; Site\_1: ClaI; Site\_2: PstI; mRNA purified from rough endoplasmic reticulum-bound polysomes was used as a template. cDNA's directionally cloned at HindIII and PstI in Bluescript SKT. Cloned into HindIII, but the site was destroyed to add an adapter sequence. So ClaI is now the restriction enzyme at site 1 of vector. Primers - 5' universal, 3' m13-20 (reverse), 3' T3, 5' T7."  
 /clone="CK02248"  
 /clone\_lib="Drosophila Embryonic Polysomal in pBS skt"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours old embryo"  
 /lab\_host="XL1 Blue MRF"  
 /lab\_host="XL1 Blue MRF"  
 complement(1..914)  
 BASE COUNT 193 a 228 c 221 g 267 t 5 others  
 ORIGIN  
 Query Match 2.8%; Score 34; DB 34; Length 914;  
 Best Local Similarity 59.7%; Pred. No. 8.73e-11;  
 Matches 105; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
 Db 480 gacacgctagcttcgttgccttaccatgctctactgtggtgatttaccatcatgtaga 539  
 QY 502 GGACATCTCTCTTTGGAGATGATGCGATGGTGTCTTGGCGCTGTATGTGCGAGCAGCA 561  
 Db 540 atgcaatggaacgagtgagatgtgacacgttcctcagttcctgtgctcatgttc 599  
 QY 562 CTCTGTGGAGTGGCGACGGCTGCTGGACCCACAGTCCAGTCTCTCTGTGGCGCTTT 621  
 Db 600 gctgtatcacgctctccagagattccgactacacgaacattggtccagatgt 655  
 QY 622 GCCTCTACGTGGGTACACCGCGTGTCTGATTACAAACACCACTGGAGCGATGT 677  
 RESULT 14  
 LOCUS AA142082 422 bp mRNA EST 25-JAN-1997  
 DEFINITION CK00216.3prime Drosophila Embryonic Polysomal in pBS skt Drosophila melanogaster cDNA clone CK00216 3'.  
 ACCESSION AA142082  
 NID 91703970  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 422)  
 AUTHORS Kopczyński, C., Serrano, T., Rubin, G. and Goodman, C.  
 TITLE mRNAs enriched for membrane and secreted proteins versus cytosolic proteins  
 JOURNAL Unpublished (1996)  
 COMMENT Other ESTs: CK00216.5prime  
 Contact: Lewis, S.  
 G. M. Rubin-Molecular and Cell Biology  
 University of Berkeley  
 539 LSA, Berkeley, CA 94720-3200, USA  
 Fax: 510 643 9947  
 Email: ES@fruitfly.berkeley.edu, http://fruitfly.berkeley.edu/  
 Insert Length: 422 Std Error: 0.00  
 Plate: 2 row: 4 column: 3  
 FEATURES  
 Location/Qualifiers  
 1..422  
 /organism="Drosophila melanogaster"  
 /note="Organ: embryo; Vector: plasmid; Site\_1: ClaI; Site\_2: PstI; mRNA purified from rough endoplasmic reticulum-bound polysomes was used as a template. cDNA's directionally cloned at HindIII and PstI in Bluescript SKT. Cloned into HindIII, but the site was destroyed to add an adapter sequence. So ClaI is now the restriction enzyme at site 1 of vector. Primers - 5' universal, 3' m13-20 (reverse), 3' T3, 5' T7."  
 /clone="CK00216"  
 /clone\_lib="Drosophila Embryonic Polysomal in pBS skt"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours old embryo"  
 /lab\_host="XL1 Blue MRF"  
 /lab\_host="XL1 Blue MRF"  
 complement(1..422)  
 BASE COUNT 140 a 91 c 102 g 89 t  
 ORIGIN  
 Query Match 2.6%; Score 32; DB 34; Length 422;  
 Best Local Similarity 60.3%; Pred. No. 7.12e-09;  
 Matches 94; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 Db 258 gacatcgacacaaatggctgttagtcggaactctgtgtgagagccgtatcaccagcga 317  
 Cp 678 GACATCGCTCCAGTGGTGTGTGTATACACACGCGGGTGTAGCCACGTAGAGGCCA 619  
 Db 318 catgagcgcaggaactggagcaggtggcagcagcattctcaactgtcttcattgtctt 377  
 Cp 618 GCCCACCAGGAAGAACTGGACTGTGGTGCAGCAGCGTGCCTTCCACACAGATCG 559  
 Db 378 acgatgtaggtaaatcaccagaatgagcagtgagta 413  
 Cp 558 TCCCTGCATACACGCCCAAGACACCATGTCAGTA 523  
 RESULT 15  
 LOCUS AA181403 446 bp mRNA EST 06-JAN-1997  
 DEFINITION zp52f10.91 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone 613099 3'.  
 ACCESSION AA181403  
 NID 91765128  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 446)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Narra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
 TITLE WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Willson RK  
 WashU-Merck EST Project

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40M13 fwd. from Amersham  
 High quality sequence stop: 286.

## FEATURES

```

source
  1..446
    /organism="Homo sapiens"
    /note="Vector: pluscript SK-; Site_1: EcoRI; Site_2:
    XhoI; Cloned unidirectionally. Primer: Oligo dt. HeLa S3
    epitheloid carcinoma cells grown to semi-confluency
    without induction. Average insert size: 1.5 kb; Uni-ZAP XR
    Vector. -5' adaptor sequence: 5' GAATCGGCACGAG 3' -3'
    adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"
    /clone="613099"
    /clone_lib="Stratagene HeLa cell s3 937216"
    /sex="female"
    /dev_stage="HeLa S3 cell line"
    /lab_host="SOLR (kanamycin resistant)"
    complement(<1..>446)
BASE COUNT      111 a      103 c      125 g      101 t      6 others
ORIGIN

```

```

Query Match      2.4%; Score 30; DB 12; Length 446;
Best Local Similarity 64.4%; Pred.No. 5.05e-07;
Matches 67; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Db 199 ttgaagaggtcagacacgaagaactatgcagcaggccaccagggctccttgagcaaat 258
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 746 TTGAAGAAGTCTGAGATGTAGCAGACAGTGAAGGGCAGCCACGATGCCCCCTGCAGGAGG 687
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 259 cctgccagaacatcactgggagtggtgtgtgtgtatcgag 302
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 686 CCACACAGGACATCGCTCCAGTGGTGTTGTATACAGACACGG 643
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: Thu Nov 6 13:05:06 1997  
 Job time : 391 secs.

**This Page Blank (uspto)**

\*\*\*\*\*  
 M P S R C H  
 (TM)  
 \*\*\*\*\*

Release 2.1d John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
 Distribution rights by IntelliGenetics, Inc.

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Nov 4 10:32:20 1997; MasPar time 6.89 Seconds  
 Tabular output not generated. 492.650 Million cell updates/sec

Title: >US-08-842-827-8  
 Description: (1-276) from US08842827.pap  
 Perfect Score: 2118

Sequence: 1 MQRWVFLDLCLVASL.....KEELEKPSLSLTILTRG 276

Scoring table: PAM 150  
 Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneses28  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21

Statistics: Mean 34.064; Variance 157.795; scale 0.216

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	99	4.7	882	9	Alpha-DNA polymerase.	1.47e+01
2	97	4.6	317	9	G-protein coupled hum	2.01e+01
3	97	4.6	317	16	G-protein coupled hum	2.01e+01
4	95	4.5	172	1	Human lymphotoxin (HL	2.73e+01
5	95	4.5	2020	2	Cardiac sodium channe	2.73e+01
6	94	4.4	154	3	Sequence of a pure po	3.18e+01
7	94	4.4	154	2	Tumoricidal polypepti	3.18e+01
8	94	4.4	154	3	Rabbit tumor necrosis	3.18e+01
9	94	4.4	154	2	Tumoricidal polypepti	3.18e+01
10	94	4.4	154	2	Rabbit tumor necrosi	3.18e+01
11	94	4.4	154	2	Tumoricidal polypepti	3.18e+01
12	94	4.4	156	8	Rabbit TNF-alpha mute	3.18e+01
13	94	4.4	234	3	Rabbit tumor necrosi	3.18e+01
14	94	4.4	235	2	Sequence of a precurs	3.18e+01
15	94	4.4	235	2	Rabbit tumor necrosi	3.18e+01
16	93	4.4	638	7	Rat brain glycine tra	3.70e+01
17	92	4.3	154	3	TNF analogue having m	4.31e+01
18	91	4.3	156	8	Bovine TNF-alpha mute	5.01e+01
19	92	4.3	157	8	Ovine TNF-alpha mutei	4.31e+01
20	92	4.3	158	4	Sequence of ovine tum	4.31e+01

21	92	4.3	233	4	R23665	Ovine TNF-alpha	4.31e+01
22	92	4.3	234	4	R22129	Sequence of ovine tum	4.31e+01
23	91	4.3	708	14	R76236	Bovine foetal heart e	5.01e+01
24	92	4.3	754	14	R76226	Bovine endothelin con	5.01e+01
25	92	4.3	882	8	R43996	Thermophilic DNA poly	4.31e+01
26	90	4.2	262	8	R42502	A. rhizogenes NIAES17	5.82e+01
27	90	4.2	279	8	R42506	Protein encoded by RO	5.82e+01
28	90	4.2	345	20	W08370	Human placenta short	5.82e+01
29	90	4.2	362	20	W08371	Human placenta long f	5.82e+01
30	89	4.2	417	7	R37684	gIV from BHV-1 strain	6.76e+01
31	89	4.2	614	19	R07908	Pemphigus vulgaris an	6.76e+01
32	89	4.2	599	6	R30742	Human pemphigus vulga	6.76e+01
33	86	4.1	69	2	P70041	Secretory signal sequ	1.05e+02
34	86	4.1	156	3	P60526	Sequence of tumour ne	1.05e+02
35	86	4.1	157	3	P60527	Sequence of tumour ne	1.05e+02
36	86	4.1	161	5	R39647	AmEPV Spheroidin asso	1.05e+02
37	87	4.1	199	21	W20506	H. pylori transmembra	9.10e+01
38	87	4.1	219	21	W20650	H. pylori surface or	9.10e+01
39	87	4.1	227	21	W09639	Human cytokine, Si2 1	9.10e+01
40	86	4.1	233	1	P90424	Human tumour necrosis	1.05e+02
41	86	4.1	235	3	P60530	Sequence of tumour ne	1.05e+02
42	87	4.1	293	21	W14481	Ramp-1.	9.10e+01
43	86	4.1	313	1	R04905	Residual protease-2	1.05e+02
44	86	4.1	313	21	W01791	Residual protease II.	1.05e+02
45	86	4.1	466	1	P96203	Human muscarinic acet	1.05e+02

#### ALIGNMENTS

RESULT 1  
 ID R45749 standard; Protein: 882 AA.  
 AC R45749;  
 DE 09-AUG-1994 (first entry)  
 DT Alpha-DNA polymerase.  
 KW Clone; DNA polymerase; hybridise; amplification; genetic engineering;  
 KW PCR; polymerase chain reaction; restriction map; plasmid.  
 OS Sulfolobus solfataricus.  
 PN J06014780-A.  
 PD 25-JAN-1994.  
 PR 30-JUN-1992; JP-194713.  
 PR 30-JUN-1992; JP-194713.  
 PA (TAKI.) TAKARA SHUZO CO LTD.  
 DR WPI; 94-061475/08.  
 DR N-PSDB; Q55724.  
 PT Cloning of alpha DNA polymerase gene - by amplifying using primer  
 PT of specified sequence and cloning to hybridise probe  
 PS Claim 2; Page 8-12; 13pp; Japanese.  
 CC The primers (Q55718-23) are used to design probes (R45746-48) for  
 CC the alpha-DNA polymerase gene (Q55724). The gene itself was  
 CC amplified using primers (Q55725-26). This method of cloning  
 CC alpha-DNA polymerase is easy and effective allowing quick and easy  
 CC production for genetic engineering research.  
 SQ Sequence 882 AA;

Query Match 4.7%; Score 99; DB 9; Length 882;  
 Best Local Similarity 42.4%; Pred. No. 1.47e+01;  
 Matches 14; Conservative 12; Mismatches 4; Indels 3; Gaps 2;

Db 605 finatvgvfgaetfypapavaesvtalgrvvi 637

QY 91 YVAAYKVVLG--TF-LFGAANSQSUTDLAKYMI 120

#### RESULT 2

ID W02661 standard; peptide: 317 AA.

AC W02661;

DT 12-NOV-1996 (first entry)

DE G-protein coupled human m2 muscarinic acetylcholine receptor.

KW G-protein coupled receptor; ligand binding assay; transmembrane domain;

KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsins;

KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;

KW odorant; cytomegalovirus; serotonergic.

OS Homo sapiens.





P50101;  
 30-OCT-1991 (first entry)  
 DE Sequence of a pure polypeptide having tumour necrosis factor  
 (TNF) activity.  
 DE Antiviral agent; lymphokine; antitumour.  
 OS Oryctolagus cuniculus.  
 PN EP-148311-A.  
 PD 17-JUL-1985.  
 PF 07-MAY-1984; 105149.  
 PR 26-DEC-1983; JP-251817.  
 PR 08-FEB-1984; JP-019850.  
 PR 18-APR-1984; JP-076584.  
 PA (ASAH) ASAHI KASEI KOGYO.  
 PI Itoh H;  
 DR WPI; 85-172835/29.  
 DR N-PSDB; N50113.  
 PT New polypeptide(s) having 154 aminoacid sequence - useful for  
 inducing necrosis of tumours and as antivirals  
 PS Claim 1; Page 54; 65pp; English.

CC The TNFs of the invention are able to induce necrosis of tumours  
 CC without toxic effects on the normal living body tissues. They are  
 CC also antivirals. Dose is 50-100 x 1,000,000 units daily for an adult.  
 CC The DNA is ligated to a replicable expression vehicle to give a  
 CC recombinant DNA. This is used to transform a microorganism,  
 CC transformants are selected and incubated to produce TNF.  
 CC Specifically claimed are plasmids pTNF-lac-1 and pTNF-lacUV5-1 and  
 CC E.coli K-12 strain transformed with them. The DNA used was  
 CC identified and isolated after studies on rabbit TNF.  
 SQ Sequence 154 AA;

Query Match 4.4%; Score 94; DB 3; Length 154;  
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;  
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

Db 31 anallangmkltdnqlvvpadglylysqvlfsgggcgsyvlthtv 77

Qy 53 THGLMA-GVITATVILVSAGEAYLVYTDRLYSRSDFNYYAAVKV 98

# RESULT 7

ID R05175 standard; protein; 154 AA.  
 AC R05175;  
 DT 18-JUL-1990 (first entry)  
 DE Tumoricidal polypeptide.  
 DE Tumoricidal polypeptide.  
 KW Tumoricidal; tumour necrosis factor.  
 PN J02019398-A.  
 PD 23-JAN-1990.  
 PF 06-JUL-1988; 166913.  
 PR 06-JUL-1988; JP-166913.  
 PA (ASAH) Asahi Chemical Ind KK.  
 DR WPI; 90-064642/09.  
 DR N-PSDB; Q02035.

PT Tumoricidal polypeptide(s) - in which DNA to code the polypeptide(s) is  
 PT constructed from human-derived TNF-coding DNA and rabbit-derived  
 PT TNF-coding DNA.  
 PS Disclosure; Page 3; 11pp; Japanese.  
 CC The sequence is derived from tumour necrosis factor DNA.  
 CC See also R05173-76 and R05189-90.  
 SQ Sequence 154 AA;

Query Match 4.4%; Score 94; DB 2; Length 154;  
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;  
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

Db 31 anallangmkltdnqlvvpadglylysqvlfsgggcgsyvlthtv 77

Qy 53 THGLMA-GVITATVILVSAGEAYLVYTDRLYSRSDFNYYAAVKV 98

# RESULT 8

ID P50138 standard; protein; 154 AA.  
 AC P50138;  
 DT 01-JAN-1980 (first entry)

DE Rabbit tumor necrosis factor.  
 KW Tumor necrosis factor; lymphokine; antitumor; ss.  
 OS Oryctolagus coturnix.  
 PN EP-146026-A.  
 PD 26-JUN-1985.  
 PF 27-NOV-1984; 114325.  
 PR 02-DEC-1983; JP-228790.  
 PR (DAIN ) DAINIPPON PHARM KK.  
 PI Yamada M, Furutani Y, Notake M;  
 DR WPI; 85-154142/26.  
 DR N-PSDB; N50196.

PT New DNA encoding rabbit tumour necrosis factor - useful in  
 PT expression of the factor in Escherichia coli for anti:tumour use  
 PS Claim 2; Page 39; 64pp; English.  
 CC The rabbit TNF can be obtained in large amounts by recombinant DNA  
 CC techniques. DNA encoding rabbit TNF is obtained by cultivation of  
 CC rabbit macrophages with inducers. A fraction containing TNF mRNA is  
 CC separated from the macrophages and used for the preparation of a  
 CC cDNA library. The TNF may be used as an antitumor agent. See also  
 CC N50197 and P50139.  
 SQ Sequence 154 AA;

Query Match 4.4%; Score 94; DB 3; Length 154;  
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;  
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

Db 31 anallangmkltdnqlvvpadglylysqvlfsgggcgsyvlthtv 77

Qy 53 THGLMA-GVITATVILVSAGEAYLVYTDRLYSRSDFNYYAAVKV 98

# RESULT 9

ID R05190 standard; protein; 154 AA.  
 AC R05190;  
 DT 18-JUL-1990 (first entry)  
 DE Tumoricidal polypeptide.  
 DE Tumoricidal polypeptide.  
 KW Tumoricidal; Tumour necrosis factor.  
 PN J02019398-A.  
 PD 23-JAN-1990.  
 PF 06-JUL-1988; 166913.  
 PR 06-JUL-1988; JP-166913.  
 PA (ASAH) Asahi Chemical Ind KK.  
 DR WPI; 90-064642/09.  
 DR N-PSDB; Q02045.

PT Tumoricidal polypeptide(s) - in which DNA to code the polypeptide(s) is  
 PT constructed from human-derived TNF-coding DNA and rabbit-derived  
 PT TNF-coding DNA.  
 PS Disclosure; Page 4; 11pp; Japanese.  
 CC The sequence is derived from tumour necrosis factor DNA.  
 CC See also R05173-76 and R05189.  
 SQ Sequence 154 AA;

Query Match 4.4%; Score 94; DB 2; Length 154;  
 Best Local Similarity 29.8%; Pred. No. 3.18e+01;  
 Matches 14; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

Db 31 anallangmkltdnqlvvpadglylysqvlfsgggcgsyvlthtv 77

Qy 53 THGLMA-GVITATVILVSAGEAYLVYTDRLYSRSDFNYYAAVKV 98

# RESULT 10

ID R03266 standard; protein; 154 AA.  
 AC R03266;  
 DT 18-JUL-1990 (first entry)  
 DE Rabbit tumour necrosis factor polypeptide.  
 DE Tumoricidal; Tumour necrosis factor.  
 KW Oryctolagus cuniculus.  
 OS Oryctolagus cuniculus.  
 PN J02019398-A.  
 PD 23-JAN-1990.  
 PF 06-JUL-1988; 166913.  
 PR 06-JUL-1988; JP-166913.  
 PA (ASAH) Asahi Chemical Ind KK.



PR 21-JUN-1985; JP-136281.  
PA (DAIN ) DAINIPPON PHARM KK.

Search completed: Tue Nov 4 10:33:59 1997  
Job time : 99 secs.

**This Page Blank (uspto)**

SUMMARIES						
Result No.	Score	Query Match	Length	DB ID	Description	Pred. No
1	145	6.8	274	11	S69561	1.31e-05
2	144	6.8	289	11	S70114	1.76e-05
3	112	5.3	317	10	D42333	1.44e-01
4	107	5.1	386	4	S2853E	5.27e-01
5	108	5.1	596	11	S46001	4.08e-01
6	105	5.0	396	4	LABERJ	8.76e-01
7	103	4.9	332	10	I39800	1.45e+00
8	103	4.9	357	9	S23349	1.45e+00
9	103	4.9	373	11	S19402	1.45e+00
10	103	4.9	417	13	S01696	1.45e+00
11	103	4.9	615	16	S69990	1.45e+00
12	103	4.9	615	11	S40624	1.45e+00
13	101	4.8	188	8	S12686	1.45e+00
14	100	4.7	381	8	S24611	2.38e+00
15	97	4.6	404	4	LABECA	3.04e+00
16	97	4.6	404	8	S21660	6.28e+00
17	97	4.6	404	8	S21660	6.28e+00
18	97	4.6	638	13	S64156	6.28e+00
19	97	4.6	638	13	I77912	6.28e+00
20	97	4.6	692	13	I57956	6.28e+00
21	95	4.5	2016	7	S48195	1.01e-01
22	95	4.5	2019	7	A33999	1.01e-01

Query Match	6.88;	Score 145;	DB 11;	Length 274;
Best Local Similarity	33.3%;	Pred. No. 1.31e-05;		
Matches	44;	Conservative 28;	Mismatches 50;	Indels 10; Gaps 10
Db	126	sinaaltgalkliignlpdvdrcipdlqkmsdsalsvfgld-ickqtnkwiylgkls	184	
QY	107	AVSQSLTDLAYMIGRLKPNFLAVCPDPSRV-NCSYVY-QLEKVCV-GNPADYTEARLS	163	
Db	185	tpsgshsfvstamftyw-q-iv-ft-trntsrsci-wcpl-lalvmmvsrvldhrhwy	238	
QY	164	FYSGHSSFGMTCMFELAYVQARCLWKWALLRPTVQFFELVAFALYGVGTRYSDYKHHWS	223	
Db	239	dvvsqavlafliv	250	
QY	224	DVLVGLQGLAV	235	
RESULT	2			
ENTRY				
TITLE		S70114	#type complete	
		hypothetical protein YDR284c	- yeast (Saccharomyces cerevisiae)	
ORGANISM			#formal_name Saccharomyces cerevisiae	
DATE			24-Aug-1996 #sequence_revision 06-Sep-1996	#text_change 06-Sep-1996

```

ACCESSIONS S70114
REFERENCE   S70114
#authors   Fulton, L.
#submission submitted to the EMBL Data Library, May 1996
#description The sequence of S. cerevisiae cosmid 9819.
#accession S70114
#molecule_type DNA
#residues_type 1-289 ##label FUL
#cross-references EMBL:U51031
GENETICS
#map_position 4R
#length 289 #molecular-weight 33514 #checksum 4703
Query Match 6.8%; Score 144; DB 11; Length 289;
Best Local Similarity 24.8%; Pred. No. 1.76e-05;
Matches 58; Conservative 65; Mismatches 92; Indels 19; Gaps 18;
Db 19 ledvf-lllmlallnyvpyqqferrfyindtishpy-a-tter-vnnmlfysfv- 73
QY 9 LLDVLCCLVASLPAITLVNAPYKRGFGYCGDSIRYPRPTITHGLMAGVTITATVIL 68
Db 74 vps-itiiligsiladrrhli-fi--lytallglslawfststfftnfknwigrldpfl 129
QY 69 VSAGEAYLVYTDRLYSRDFNNYAAVYKVLGTLFGAAVSQSLDLAKYMGRLKPNFL 128
Db 130 drcqpveglpid-tlftakd-vcttknherlldgfttp-sghsaeefaglyfwlwg 186
QY 129 AVCDB-DMSRVNCSYVQLEKVCNGPAD-VTEA-RLSYSGHSSFGMYCMVFLALYVQA 185
Db 187 qltspmpirwkmavflpilgaal-ialstrtdgyrhfdvdlgsmlygima 239
QY 186 RLCWKWARL-L-RPTVQFF-LVAFALYVGTVRSYDKHHWSDVLYLGLGALVA 235
RESULT 3
ENTRY     D64233 #type complete
TITLE     hypothetical protein MG302 - Mycoplasma genitalium (SGC3)
ORGANISM  #formal_name Mycoplasma genitalium
DATE      10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
ACCESSIONS D64233
REFERENCE   D64233
#authors   Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
           R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
           Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
           Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
           Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
           J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
           Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
           C.A.; Venter, J.C.
#journal   Science (1995) 270:397-403
#title     The minimal gene complement of Mycoplasma genitalium.
#accession D64233
#status    preliminary
#molecule_type DNA
#residues  1-317 ##label TIGR
#cross-references GB:L43967; TIGR:MG302
#experimental_source strain G-37
#note      neither nucleotide sequence nor conceptual translation
           is shown
GENETICS
#genetic_code SGC3
SUMMARY
#length 317 #molecular-weight 36667 #checksum 8963
Query Match 5.3%; Score 112; DB 10; Length 317;
Best Local Similarity 24.5%; Pred. No. 1.44e-01;
Matches 25; Conservative 27; Mismatches 4; Indels 3; Gaps 3;
Db 167 lslllavikflptvkselirkqaqatrgfylnkcsfnlnpfkiktlfipvllstvkke 226
QY 6 VFVLDVLCCLVASLPAITLVNAPYKRGFGYCGDSIRYPRPTI-THGLMAGVTITA 64
Db 227 ttafalgakgdylnntnrthypkyn-llngvfllyglilis 267

```

```

QY 55 TVILVSAGEAY-LVYTDRLYSRDFNNYAAVYKVLGTLFEG 105
RESULT 4
ENTRY     S2B5SE #type complete
TITLE     stage V sporulation protein E - Bacillus subtilis
ALTERNATE_NAMES spoVE protein
ORGANISM   #formal_name Bacillus subtilis
DATE       28-Dec-1987 #sequence_revision 06-Jan-1995 #text_change
ACCESSIONS S10243; A29756; E47691; S22211
REFERENCE   S10243
#authors   Sato, T.; Theeragool, G.; Yamamoto, T.; Okamoto, M.;
           Kobayashi, Y.
#journal   Nucleic Acids Res. (1990) 18:4021
#title     Revised nucleotide sequence of the sporulation gene spoVE
           from Bacillus subtilis.
#cross-references MUID:90326557
#accession S10243
#status    translation not shown
#molecule_type DNA
#residues  1-366 ##label SAT
#cross-references EMBL:X51419
REFERENCE   A29756
#authors   Bugaichuk, U.D.; Piggot, P.J.
#journal   J. Gen. Microbiol. (1986) 132:1883-1890
#title     Nucleotide sequence of the Bacillus subtilis developmental
           gene spoVE.
#cross-references MUID:87085411
#accession A29756
#molecule_type DNA
#residues  'MVQIADGYLLFPS', 87-366 ##label BUG
REFERENCE   A47691
#authors   Daniel, R.A.; Errington, J.
#journal   J. Gen. Microbiol. (1993) 139:361-370
#title     DNA sequence of the mreE-murD region of Bacillus subtilis
           168.
#cross-references MUID:93171879
#contents  168
#accession E47691
#molecule_type DNA
#residues  1-40 ##label DAN
#cross-references NCBI:125659; NCBI:125665
#note      sequence extracted from NCBI backbone
REFERENCE   S22211
#authors   Henriques, D.O.; de Lencastre, H.; Piggot, P.J.
#submission submitted to the EMBL Data Library, January 1992
#accession S22211
#molecule_type DNA
#residues  343-366 ##label HEN
#cross-references EMBL:X64259
GENETICS
#gene      spoVE
#map_position 133 (degrees)
#start_codon TTG
CLASSIFICATION #superfamily rod shape-determining protein
KEYWORDS      sporulation
SUMMARY       #length 366 #molecular-weight 40132 #checksum 1239
Query Match 5.1%; Score 107; DB 4; Length 366;
Best Local Similarity 22.6%; Pred. No. 5.27e-01;
Matches 19; Conservative 25; Mismatches 35; Indels 5; Gaps 5;
Db 53 agigviam-fimnvdwt-wrtwskllm-vicfillvllpvgvmrngerawigvga 109
QY 169 SSFGMYCMVFLALYVQARLCWK-WARLRTVVOFFLVAFALYGVYTRVSDYKHHWSDVLY 227
Db 110 fsiqpsefkmklamlaflakflsek 133
QY 228 GLLGALVAALT-VCIYSDFPKAR 250

```



```

polymorphism.
#cross-references MWID:92201631
#accession S2f349
#status preliminary
#molecule_type DNA
##residues 1-367 #label WAN
##cross-references EMBL:X60666
##note the source is designated as Salmonella enterica
SUMMARY #length 367 #molecular-weight 42049 #checksum 183

Query Match 4.9%; Score 103; DB 9; Length 367;
Best Local Similarity 18.8%; Pred. No. 1.45e+00;
Matches 15; Conservative 32; Mismatches 30; Indels 3; Gaps 3;

Db 155 vsnallsvaisvilitvaitsyls-pqyilvk-dyklnficqlvtgmsfyiqyvirn 212
      ::::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 52 ITHGLMAGVTITATVILVSAGEALVYTDRLSKSDFN-NYVAAYVKVLGTLFLFGAASQ 110
      ::::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 213 qiyvnlfnfyfilltviilyv 232
      ::::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 111 SLTDLAKYMIGRLKNFLAV 130
      ::::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 9
ENTRY #S19402 #type complete
TITLE hypothetical protein YCL070c - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Feb-1995
ACCESSIONS S19402
REFERENCE S19400
#authors van der Aart, Q.J.M.; Steensma, H.Y.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19402
#molecule_type DNA
##residues 1-373 #label AAR
##cross-references EMBL:X59720
GENETICS
#map_position 3L
SUMMARY #length 373 #molecular-weight 42503 #checksum 4433

Query Match 4.9%; Score 103; DB 11; Length 373;
Best Local Similarity 23.5%; Pred. No. 1.45e+00;
Matches 23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db 47 elmsacfdalkykl-l-listafvcgfgisldylrstlygyatnsysehsllstqvina 105
      ::: | | | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 11 DVLCLLVASLPAILTLNYPARKGF-YCGDGSIR--YP-YRPDTIT-HGLMAGVTITAT 65
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 106 vswagsgqvysrlsdh-fgrlrif-lvatifyimgtl 141
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 66 VILVSAGEALVYTDRLSRSDENNYYAAVVKVLGTFL 103
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
ENTRY #S01696 #type complete
TITLE gene P3 protein - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Jun-1993
ACCESSIONS S01696
REFERENCE S01696
#authors Alcalay, M.; Toniolo, D.
#journal Nucleic Acids Res. (1988) 16:9527-9543
#title CpG Islands of the X chromosome are gene associated.
#cross-references MWID:89041548
#accession S01696
#molecule_type DNA
##residues 1-477 #label ALC
##cross-references EMBL:X12458
SUMMARY #length 477 #molecular-weight 50332 #checksum 7015

```



```

Db 106 vvsqgvyvrsldh-fgrlrif-lvatifimgtll 141
QY 66 VILVSAGEALVYTDRLYSRDFNNYVAAYKVLGTFL 103

RESULT 13
ENTRY #type complete
TITLE coat protein - belladonna mottle virus
ALTERNATE_NAMES virion protein
ORGANISM #formal_name belladonna mottle virus
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
31-Dec-1993
ACCESSIONS S12686
REFERENCE #authors Ding, S.W.; Howe, J.; Mackenzie, A.; Skotnicki, M.; Gibbs, A.
#journal Nucleic Acids Res. (1990) 18:6138
#title Nucleotide sequence of the virion protein gene of belladonna
mottle tymovirus.
#cross-references MUID:91045090
#accession S12686
#molecule_type DNA
#residues 1-188 #label DIN
#cross-references EMBL:X54529
#note the authors translated the codon ACC for residue 64 as U
and ACA for residue 167 as U
CLASSIFICATION #superfamily turnip yellow mosaic virus coat protein
KEYWORDS #length 188 #molecular-weight 19891 #checksum 5832
FEATURE 25-45
51-71 #domain transmembrane #status predicted #label TM1\
76-97 #domain transmembrane #status predicted #label TM2\
104-124 #domain transmembrane #status predicted #label TM3\
140-160 #domain transmembrane #status predicted #label TM4\
166-186 #domain transmembrane #status predicted #label TM6\
256-332 #region 11-residue repeats (P-D-N-P-D-N-G-P-Q-D)
SUMMARY #length 404 #molecular-weight 43769 #checksum 7535

Query Match 4.8%; Score 101; DB 8; Length 188;
Best Local Similarity 24.1%; Pred. No. 2.38e+00;
Matches 20; Conservative 28; Mismatches 30; Indels 5; Gaps 5;

Db 37 qfeatsvgvyetlaq-v-nlssdsiakltsgyrrakvvelftitptra-idcpvtvd 93
QY 87 DENNYAAVYKVLGTFLFGAAYQSISLTLAK-YMIGRLKPNFLAVCPDPSRVNCSYVQ 145
Db 94 vavv-panstappskilslyggq 115
QY 146 LEKVRGNPADVTPEARLSFYSGH 168

RESULT 14
ENTRY #type complete
TITLE latent membrane protein - human herpesvirus 4
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
12-Apr-1995
ACCESSIONS S24611
REFERENCE #authors Chang, Y.S.
#submission submitted to the EMBL Data Library, June 1992
#accession S24611
#status preliminary
#molecule_type DNA
#residues 1-381 #label CHA
#cross-references EMBL:X68863
GENETICS
#introns 90/1; 119/1
CLASSIFICATION #superfamily Epstein-Barr virus latent membrane protein
SUMMARY #length 381 #molecular-weight 41372 #checksum 3544

Query Match 4.7%; Score 100; DB 8; Length 381;
Best Local Similarity 25.7%; Pred. No. 3.04e+00;
Matches 18; Conservative 21; Mismatches 27; Indels 4; Gaps 4;

Db 104 alylgivlf-ifgcilvfgiwyfleil-wrlgatlwqlafil-afflaillialylq 161
QY 163 SFYSGHSSFGMY-CM-VF-LALYVQARLCWKARLLRPTVQFFLVAFALYGVTRVSDY-KH 220
Db 162 nwtwllvdl 171
QY 221 HWSDLVGLL 230

RESULT 15
ENTRY #type complete
TITLE latent membrane protein - human herpesvirus 4 (strain CAO)
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
09-Sep-1994
ACCESSIONS J01434
REFERENCE #authors Hu, L.F.; Zabarovsky, E.R.; Chen, F.; Cao, S.L.; Ernberg, I.;
Klein, G.; Winberg, G.
#journal J. Gen. Virol. (1991) 72:2399-2409
#title Isolation and sequencing of the Epstein-Barr virus BNLF-1
gene (LMP1) from a Chinese nasopharyngeal carcinoma.
#cross-references MUID:92013956
#accession J01434
#molecule_type DNA
#residues 1-404 #label HUL
#note the authors translated the codon AAA for residue 358 as
Ala
GENETICS
#gene LMP1; BNLF-1
#introns 90/1; 119/1
CLASSIFICATION #superfamily Epstein-Barr virus latent membrane protein
KEYWORDS transmembrane protein
FEATURE 25-45
51-71 #domain transmembrane #status predicted #label TM1\
76-97 #domain transmembrane #status predicted #label TM2\
104-124 #domain transmembrane #status predicted #label TM3\
140-160 #domain transmembrane #status predicted #label TM4\
166-186 #domain transmembrane #status predicted #label TM6\
256-332 #region 11-residue repeats (P-D-N-P-D-N-G-P-Q-D)
SUMMARY #length 404 #molecular-weight 43769 #checksum 7535

Query Match 4.6%; Score 97; DB 4; Length 404;
Best Local Similarity 29.2%; Pred. No. 6.28e+00;
Matches 21; Conservative 18; Mismatches 26; Indels 7; Gaps 7;

Db 104 alylgivlf-ifgcilvfgiwyfleil-wrlgatlwqlafil-afflaillialyl 160
QY 163 SFYSGHSSFGMY-CM-VF-LALYVQARLCWKARLLRPTVQFFLVAFALYGVTRVSDY- 218
Db 161 qqnwtwllvdl 172
QY 219 KHRWSDVLVGLL 230

Search completed: Tue Nov 4 10:32:01 1997
Job time : 77 secs.

```

**This Page Blank (uspto)**

\*\*\*\*\*  
MORSE (TM)  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Nov 4 10:29:23 1997; MasPar time 7.58 Seconds  
Tabular output not generated.  
772.348 Million cell updates/sec

Title: >US-08-842-827-8  
Description: (1-276) from US08842827.ppt  
Perfect Score: 2118  
Sequence: 1 MORRWVFLVDVLCVLSL.....KEEELERKPSLSLTILTRG 276

Scoring table: PAM 150  
Gap 11  
Searched: 59021 seqs, 21210388 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11  
Statistics: Mean 48.002; Variance 91.621; scale 0.524

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	461	21.8	341	11	YSX3_CAEEL	HYPOTHETICAL 39.0 KD	2.95e-67
2	112	5.3	317	10	Y302_MYCGE	HYPOTHETICAL PROTEIN	1.65e-02
3	107	5.1	366	9	SPSE_BACSU	STAGE V SPOULATION P	7.74e-02
4	108	5.1	596	11	YBX2_YEAST	HYPOTHETICAL AMINO-AC	5.70e-02
5	105	5.0	386	6	LMPL_EBRV	LATENT MEMBRANE PROTE	1.42e-01
6	103	4.9	332	2	CCPA_BACME	GLUCOSE-RESISTANCE AM	2.57e-01
7	103	4.9	477	7	P3_HUMAN	P3 PROTEIN	2.57e-01
8	103	4.9	615	11	YCHO_YEAST	HYPOTHETICAL 68.9 KD	2.57e-01
9	103	4.9	615	11	YK86_YEAST	HYPOTHETICAL 69.0 KD	2.57e-01
10	99	4.7	309	11	YQGH_BACSU	PROBABLE ABC TRANSPOR	8.29e-01
11	98	4.6	261	11	YV23_MYCLE	HYPOTHETICAL 27.9 KD	1.10e-00
12	97	4.6	404	6	LMPL_EBVC	LATENT MEMBRANE PROTE	1.47e-00
13	97	4.6	616	11	YGO2_YEAST	HYPOTHETICAL 72.6 KD	1.47e-00
14	97	4.6	692	6	NTGL_HUMAN	SODIUM- AND CHLORIDE-	1.47e-00
15	95	4.5	2019	2	CIN5_RAT	SODIUM CHANNEL PROTEI	2.57e-00
16	94	4.4	233	9	TNFA_BOVIN	TUMOR NECROSIS FACTOR	3.39e-00
17	94	4.4	235	9	TNFA_RABIT	TUMOR NECROSIS FACTOR	3.39e-00
18	94	4.4	334	2	CCPA_BACSU	GLUCOSE-RESISTANCE AM	3.39e-00
19	94	4.4	386	6	LMPL_EBV	LATENT MEMBRANE PROTE	3.39e-00
20	94	4.4	490	1	ACM4_CHICK	MUSCARINIC ACETYLCHOL	3.39e-00
21	94	4.4	555	6	NRM1_CHICK	NATURAL RESISTANCE-AS	3.39e-00
22	93	4.4	633	6	NTGL_RAT	SODIUM- AND CHLORIDE-	4.47e-00

ALIGNMENTS									
RESULT 1		STANDARD:		PRT:		341 AA.			
ID	YSX3_CAEEL								
AC	Q10022;								
DT	01-OCT-1996 (REL. 34, CREATED)								
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)								
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)								
DE	HYPOTHETICAL 39.0 KD PROTEIN T28D9.3 IN CHROMOSOME II.								
GN	T28D9.3.								
OS	CAENORHABDITIS ELEGANS.								
OC	EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN-BRISTOL N2;								
RA	WATERSTON R.;								
RL	SUBMITTED (JUN-1995) TO EMEL/GENBANK/DBJ DATA BANKS.								
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).								
CC	-1- SIMILARITY: WEAK, TO YEAST D9719.9.								
DR	EMBL; U28738; G861267; -								
DR	WORMPEP; T28D9.3; CE02068.								
KW	HYPOTHETICAL PROTEIN; TRANSMEMBRANE.								
FT	TRANSMEM 30 50								
FT	TRANSMEM 71 91								
FT	TRANSMEM 122 142								
FT	TRANSMEM 223 243								
FT	TRANSMEM 257 277								
SQ	SEQUENCE 341 AA; 39028 MW; 89AE6E81 CRC32;								
Query Match 21.8%; Score 461; DB 11; Length 341;									
Best Local Similarity 35.1%; Pred. No. 2.95e-67;									
Matches 93; Conservative 63; Mismatches 94; Indels 15; Gaps 14;									
Db	21	grskqfgislfiffilaataavtvtllgvs-qrgffcdsdsiryeakdtitavqlmly	79						
QY	2	QRWVF-VLLDVLCVLSLPAITLVNAPYKRGVCGDDSIYRYPRTITHG-LMA- 58							
Db	80	nlvlnaatvlfveyyrmqkvesnnpvyrnrnhlflvrllytyfgsqigfvmial	139						
QY	59	GVVI-TATVILVSAGEALVYTDRLYSRSDF-NNVAAVY-KVLGTFLGA-A-VSQ-SL	112						
Db	140	nivtkhvgvrlphrfdvcklandtcvtdgshryit-dyctgppelvlearksfysghs	198						
QY	113	TDLAKNIGRLKPNFLAVCD-PDWSRV--NCVVYQLEKVCVGRGNPADVTEARLSFYSGHS	169						
Db	199	avslcyatwsalyiqarlpvlnnrivvpisqtlmfmglgistsritdnkhwsdvlvg	258						
QY	170	SEGYCMVFLALVQARLCKWA-RLLRPTVQFLVAPALVGVYTRVSDYKHEWSVDVVG	228						

Db 259 ifigiflavvtctftwtdlfsnnste 283  
 QY 229 LQGGALVAALVCISDFFFRARPPQ 253

RESULT 2  
 ID Y302\_MYCGE STANDARD; PRT; 317 AA.  
 AC P47544;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL PROTEIN MG302.  
 GN MG302.  
 OS MYCOPLASMA GENITALIUM.  
 OC PROKARYOTA; TENERICUTES; MOLLICUTES; MYCOPLASMA; MYCOPLASMATALES;  
 OC MYCOPLASMATACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33530 / G-37;  
 RX MEDLINE; 96026346.  
 RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,  
 RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,  
 RA FRITCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,  
 RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,  
 RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,  
 RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;  
 SCIENCE 270:397-403(1995).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 DR EMBL; U39711; G1046001;  
 KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
 FT TRANSMEM 18 38 POTENTIAL.  
 FT TRANSMEM 58 78 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 130 150 POTENTIAL.  
 FT TRANSMEM 159 179 POTENTIAL.  
 FT TRANSMEM 202 222 POTENTIAL.  
 FT TRANSMEM 252 272 POTENTIAL.  
 SQ SEQUENCE 317 AA; 36667 MW; EF5F2980 CRC32;

Query Match 5.3%; Score 112; DB 10; Length 317;  
 Best Local Similarity 24.5%; Pred. No. 1.65e-02;  
 Matches 25; Conservative 27; Mismatches 47; Indels 3; Gaps 3;  
 Db 167 lsllavfkilptvkselirkqagatrflynkcsfnlpfkixtlfipvllsvkkte 226  
 QY 6 VFVLDVLCILVASLPFAILLVNAPYKRGFCGDDSIKPYRPTTI-THGLMAGVTITA 64  
 Db 227 ttafalqkqydlntnrthypkyn-llngvflvglllfs 267  
 QY 65 TWIIYSAGEAY-LVYTDRLYSRDFNNYVAAYKVLGTFLFG 105

RESULT 3  
 ID SP5E\_BACSU STANDARD; PRT; 366 AA.  
 AC P07373;  
 DT 01-APR-1988 (REL. 07, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE STAGE V SPOULATION PROTEIN E.  
 GN SPOVE.  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 80376557  
 RA SATO T., THERAGOOOL G., YAMAMOTO T., OKAMOTO M., KOBAYASHI Y.;  
 RL NUCLEIC ACIDS RES. 18:4021-4021(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87085411.  
 RA BUGAICHUK U.D., PIGGOT P.J.;

J. GEN. MICROBIOL. 132:1883-1890(1986).  
 [3]  
 RN SEQUENCE OF 1-46 FROM N.A.  
 RA THERAGOOOL G., MIYAO A., YAMADA K., SATO T., KOBAYASHI Y.;  
 RL SUBMITTED (JAN-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.  
 [4]  
 RN SEQUENCE OF 1-40 FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE; 93171879  
 RA DANIEL R.A., ERRINGTON J.;  
 RL J. GEN. MICROBIOL. 139:361-370(1993).  
 CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE NOT ONLY DURING SPOULATION,  
 CC BUT ALSO DURING VEGETATIVE GROWTH.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- CAUTION: BELONGS TO THE FTSW/RODA/SPOVE FAMILY.  
 CC -!- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINUS  
 CC AND IS SHORTER DUE TO A FRAMESHIFT.  
 DR EMBL; X51419; G580937;  
 DR EMBL; M15742; G143657; ALT\_FRAME.  
 DR EMBL; D14109; G216343;  
 DR EMBL; Z15056; ; NOT\_ANNOTATED\_CDS.  
 DR PIR; A29756; S2BSSE.  
 DR PIR; S10243; S10243.  
 DR SUBTILIS; BG10226; SPOVE.  
 DR PROSITE; PS00428; FTSW\_RODA\_SPOVE.  
 KW SPOULATION; CELL SHAPE; TRANSMEMBRANE.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 77 97 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 144 164 POTENTIAL.  
 FT TRANSMEM 185 205 POTENTIAL.  
 FT TRANSMEM 227 247 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 306 326 POTENTIAL.  
 FT TRANSMEM 337 357 POTENTIAL.  
 SQ SEQUENCE 366 AA; 40132 MW; ECC56F98 CRC32;  
 Query Match 5.1%; Score 107; DB 9; Length 366;  
 Best Local Similarity 22.6%; Pred. No. 7.74e-02;  
 Matches 19; Conservative 25; Mismatches 55; Indels 5; Gaps 5;  
 Db 53 agigviam-ffimvdywt-wrtwskllm-vicffllvlpvgvmvngsrswivgva 109  
 QY 169 SSFGMYCNVFLALYVQARLCWK-WARLLRPTVQFFLVAFLYVGYTRVSDYKHHWSDLV 227  
 Db 110 fsiqpsfemklamlaflakflsek 133  
 QY 228 GLQGLVAALT-VCIISDFFKAR 250  
 RESULT 4  
 ID YBY2\_YEAST STANDARD; PRT; 596 AA.  
 AC P38090;  
 DT 01-OCT-1994 (REL. 30, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL AMINO-ACID PERMEASE IN YMA2-CKS1 INTERGENIC REGION.  
 GN YBR132C OR YBR1007.  
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE; 94378717  
 RA BECAM A.-M., CULLIN C., GRZYBOWSKA E., LACHROUTE F., NASR F.,  
 RA OZIER-KALOGEROPOULOS O., PALUCHA A., SLONIMSKI P.P., ZAGULSKI M.,  
 RA HERBERT C.J.;  
 RL YEAST 10:S1-s11(1994).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE; 95042830.  
 RA NASR F., BECAM A.-M., GRZYBOWSKA E., ZAGULSKI M., SLONIMSKI P.P.,



CC CODES FOR A PROTEIN WITH HOUSEKEEPING FUNCTIONS.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC - SIMILARITY: TO P3 PROTEIN OF ANIMALS AND YEASTS.  
 CC - SIMILARITY: BELONGS TO THE SODIUM:BILE ACID SYMPORTER FAMILY (SBF).

DR EMBL: X12458; G35188; -.  
 DR EMBL: L41140; G1203976; -.  
 DR PIR: S01696; S01696.  
 DR MIM: 312090; -.

KW TRANSMEMBRANE; TRANSPORT; SYMPORT.  
 SQ SEQUENCE 477 AA; 50332 MW; 65304BA7 CRC32;

Query Match 4.9%; Score 103; DB 7; Length 477;

Best Local Similarity 26.5%; Pred. No. 2.57e-01;

Matches 22; Conservative 30; Mismatches 25; Indels 6; Gaps 6;

Db 276 lligdvtlaismftlstvaatgflpissalsyrlslhetlvpisklgtllf-iaip1 334

QY 56 LMAQ-VTITAV-ILVS-AGEAYLVYDRLYSR-DFNNYV-AAVKYVLGTFLEGAAYV 110

Db 335 avglvlskpkfsgllqvvpk 357

QY 111 SLTDLAKYMGRLKPNFLAVGDP 133

# RESULT 8

ID YCHO\_YEAST STANDARD; PRT; 615 AA.

AC P25597; P25599;

DT 01-MAY-1992 (REL. 22, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 68.9 KD PROTEIN IN HML 5' REGION.

GN YCL070C/71C/73C.

OS SACSCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMICETES.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DER AART Q.J.M., STEENSA H.Y.;

RL SUBMITTED (MAR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP REVISIONS.

RA GROMADKA R.;

RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC - SIMILARITY: BELONGS TO THE YCL70C/YHL047C/YKRI06W FAMILY.

DR EMBL: X59720; E264591; -.

DR PIR: S19402; S19402.

DR PIR: S19403; S19403.

DR PIR: S19405; S19405.

KW HYPOTHETICAL PROTEIN: TRANSMEMBRANE.

FT TRANSMEM 59 79 POTENTIAL.

FT TRANSMEM 121 141 POTENTIAL.

FT TRANSMEM 153 173 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 276 296 POTENTIAL.

FT TRANSMEM 308 328 POTENTIAL.

FT TRANSMEM 344 364 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 409 429 POTENTIAL.

FT TRANSMEM 441 461 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT TRANSMEM 548 568 POTENTIAL.

SQ SEQUENCE 615 AA; 68898 MW; 0BB5B37 CRC32;

Query Match

Best Local Similarity 4.9%; Score 103; DB 11; Length 615;

Matches 23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db 47 elmsaqfoslkykil-listafvcgfgisldylrstygtatnsysehslstqvina 105

QY 11 DVLCLLVASLPFAILLVNAPYKRGF-YCGDDSI--YP-YRPDTIT-HGLMAGVTITAT 65

Db 106 vvsygsqvvsrlsdh-fgrlrlf-lvatifyimgtli 141

QY 66 VILVSAGEAYLVYDRLYSRDNFNYYAAVYKVLGTF 103

## RESULT 9

ID YK86\_YEAST STANDARD; PRT; 615 AA.

AC P36173;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 69.0 KD PROTEIN IN SIR1 3' REGION.

GN YKRI06W.

OS SACSCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMICETES.

RN [1]

RP SEQUENCE FROM N.A.

RA GAILLON L., DUJON B.;

RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC - SIMILARITY: BELONGS TO THE YCL70C/YHL047C/YKRI06W FAMILY.

DR EMBL: Z28202; G486617; -.

DR PIR: S40624; S40624.

KW HYPOTHETICAL PROTEIN: TRANSMEMBRANE.

FT TRANSMEM 59 79 POTENTIAL.

FT TRANSMEM 121 141 POTENTIAL.

FT TRANSMEM 153 173 POTENTIAL.

FT TRANSMEM 217 237 POTENTIAL.

FT TRANSMEM 276 296 POTENTIAL.

FT TRANSMEM 308 328 POTENTIAL.

FT TRANSMEM 344 364 POTENTIAL.

FT TRANSMEM 384 404 POTENTIAL.

FT TRANSMEM 408 428 POTENTIAL.

FT TRANSMEM 441 461 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT TRANSMEM 549 569 POTENTIAL.

SQ SEQUENCE 615 AA; 68954 MW; A8423CFB CRC32;

Query Match

Best Local Similarity 4.9%; Score 103; DB 11; Length 615;

Matches 23; Conservative 34; Mismatches 33; Indels 8; Gaps 7;

Db 47 elmsaqfoslkykil-listafvcgfgisldylrstygtatnsysehslstqvina 105

QY 11 DVLCLLVASLPFAILLVNAPYKRGF-YCGDDSI--YP-YRPDTIT-HGLMAGVTITAT 65

Db 106 vvsygsqvvsrlsdh-fgrlrlf-lvatifyimgtli 141

QY 66 VILVSAGEAYLVYDRLYSRDNFNYYAAVYKVLGTF 103

## RESULT 10

ID YQGH\_BACSU STANDARD; PRT; 309 AA.

AC P46339;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE PROBABLE ABC TRANSPORTER PERWEASE PROTEIN IN SODA-COMGA INTERGENIC REGION (ORF72).

GN YQGH.

OS BACILLUS SUBTILIS.

OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168 / JH642;

RA TAKEMARU K., MIZUNO M., SATO T., TAKEUCHI M., KOBAYASHI Y.;

RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168 / JH642;

RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEMARU K., HOSONO S.;

RA SATO T., TAKEUCHI M.;

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

```

CC -1- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM.
CC PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE SUBSTRATE ACROSS
CC THE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYST
CC SUBFAMILY.
DR EMBL; D58414; G903304; -
DR EMBL; D84432; G1303855; -
DR SUBTILIST; BG11376; YOGH.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR.
KW HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE.
FT TRANSMEM 30 50
FT TRANSMEM 88 108
FT TRANSMEM 133 153
FT TRANSMEM 165 185
FT TRANSMEM 214 234
FT TRANSMEM 280 300
FT TRANSMEM 309 AA; 33219 MW; ED7700D9 CRC32;
SQ SEQUENCE 309 AA; 33219 MW; ED7700D9 CRC32;

Query Match 4.7%; Score 99; DB 11; Length 309;
Best Local Similarity 24.1%; Pred. No. 8.29e+01;
Matches 21; Conservative 24; Mismatches 36; Indels 6; Gaps 6;

Db 38 imlaasvatiflgvkglsflvngvspiefltslnwptdsdkpygl-pfifgsfvt 96
Qy 56 LMAGVTITATVIL-VSAGEAYLYV-TDRL-YKRS-DFNNYVAV-YKVLGTFLFGAAVSQ 110
Db 97 lsallaaplgtagpifmtelapnwgk 123
Qy 111 SLTDLAKYIMIGRLKPNFLAVCDPWSR 137

RESULT 11
ID YV23_MYCLE STANDARD; PRT; 261 AA.
AC P54580;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 27.9 KD PROTEIN B2168_C2_209.
GN B2168.C2_209.
OS MYCOBACTERIUM LEPRAE.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH D.R., ROBISON K.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO M.TUBERCULOSIS MTCY20G9.23.
DR EMBL; U00018; G467047; -
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 181 201
FT TRANSMEM 207 227
FT TRANSMEM 236 256
FT TRANSMEM 261 AA; 27924 MW; 9712F280 CRC32;
SQ SEQUENCE 261 AA; 27924 MW; 9712F280 CRC32;

Query Match 4.6%; Score 98; DB 11; Length 261;
Best Local Similarity 31.7%; Pred. No. 1.10e+00;
Matches 20; Conservative 16; Mismatches 22; Indels 5; Gaps 5;

Db 191 vvfaglfvafdq-lwrnsiv-avlsavlvgllvvg-vrvvrtedstliavvga 247
Qy 176 MVFLA-LYVQ-ARLCKWALLRLPTVQFFLVAFALVGYTRVSDYKHHWSDVLVGLQGA 233
Db 248 lit 250
Qy 234 LVA 236

RESULT 12
ID LMP1_EBVC
AC P29362;
DT 01-DEC-1992 (REL. 24, CREATED)

```

```

DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE LATENT MEMBRANE PROTEIN 1 (LMP-1) (P63).
GN BNLF1.
OS EPSTEIN-BARR VIRUS (STRAIN CAO) (HUMAN HERPESVIRUS 4).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 92013956.
RA HU L.F., ZABAROVSKY E.R., CHEN F., CAO S.L., ERNBERG I., KLEIN G.,
RA WINBERG G.;
RL J. GEN. VIROL. 72:2399-2409(1991).
CC -1- FUNCTION: THE LATENT MEMBRANE PROTEIN HAS TRANSFORMING ACTIVITY.
DR EMBL; X58140; G22938; -
DR PIR; J01434; LABECA.
DR PIR; S21660; S21660.
KW TRANSMEMBRANE; PHOSPHORYLATION; TRANSFORMING PROTEIN.
FT DOMAIN 1 24
FT TRANSMEM 25 44
FT TRANSMEM 52 72
FT TRANSMEM 77 97
FT TRANSMEM 105 125
FT TRANSMEM 139 159
FT TRANSMEM 166 186
FT TRANSMEM 187 404
FT DOMAIN 187 404
FT TRANSMEM 187 404
SQ SEQUENCE 404 AA; 43769 MW; 154E84C3 CRC32;

Query Match 4.6%; Score 97; DB 6; Length 404;
Best Local Similarity 29.2%; Pred. No. 1.47e+00;
Matches 21; Conservative 18; Mismatches 26; Indels 7; Gaps 7;

Db 104 alylgivlf-ifgcilvfgiwyfleiil-wrlgatllqlllefl-afflaifillilaly 160
Qy 163 SFYSGHSSFGMT-CM-VF-LALYVQARLCWRLRLPTVQFFLVAFALVGYTRVSDY- 218
Db 161 qdpwtllvdl 172
Qy 219 KHWSDVLVGLL 230

RESULT 13
ID YG02_YEAST STANDARD; PRT; 616 AA.
AC P30777;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 72.6 KD PROTEIN IN MRF1-SEC27 INTERGENIC REGION.
GN YGL142C.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMICOTINA; HEMIASCOMICETES.
RN [1]
RP SEQUENCE FROM N.A.
RA VOLCKAERT G., VOET M., VERHASSELT P., DEFOOR E.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 339-616 FROM N.A.
RX MEDLINE; 93117110.
RA PEL H.J., MAAT M.J., REP M., GRIVELL L.A.;
RL NUCLEIC ACIDS RES. 20:6339-6346(1992).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: SOME, TO S.POMBE SPAC2G11.09 AND SPAC468.12C.
DR EMBL; 272664; E243764; -
DR EMBL; X60381; G388249; -
DR PIR; S28601; S28601.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 17 37
FT TRANSMEM 87 107
FT TRANSMEM 137 157
FT TRANSMEM 189 209
FT TRANSMEM 241 261
FT TRANSMEM 279 299
FT TRANSMEM 339 359
FT TRANSMEM 393 413

```







**This Page Blank (uspto)**